

Pint-sized proteins pack
a powerful punch p. 296

A technology roadmap for
writing genomes p. 310

The geochemistry
of exoplanets p. 356

Science

\$15
18 OCTOBER 2019
sciencemag.org

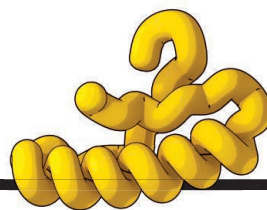
AAAS

SHAPE SHIFTER

Colonies of microbial eukaryotic cells display
coordinated contractility pp. 300 & 326



CONTENTS



296

Small proteins
flex their muscle

18 OCTOBER 2019 • VOLUME 366 • ISSUE 6463



NEWS

IN BRIEF

286 News at a glance

IN DEPTH

289 Drought test begins in Biosphere 2 rainforest

Tracing carbon in closed ecosystem could reveal tree and soil response to water stress *By E. Stokstad*

290 NIH 'high risk, high reward' awardees skew male—again

Agency struggles to attract female applicants, keep review process gender neutral *By M. Wadman*

291 Sauropods get a new diet and a new look

The behemoths may have chomped nutritious horsetails—with the help of sturdy beaks *By J. Pickrell*

292 Lithium-ion battery development takes Nobel

Three chemists honored for creating powerful, light, rechargeable cells *By R. F. Service*

293 Rival theories face off over brain's source of consciousness

\$20 million project puts competing ideas to the test *By S. Reardon*
PODCAST

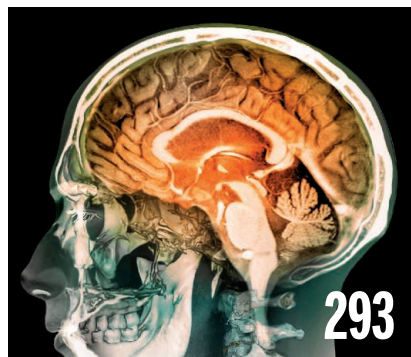
294 Texas cancer agency seeks new vote of approval

Ballot initiative for CPRIT would raise another \$3 billion to fuel state's cancer research *By J. Kaiser*

FEATURES

296 Outsize impact

Pint-size proteins, found everywhere from venom to muscles, can block or boost their bigger kin *By M. Leslie*



INSIGHTS

PERSPECTIVES

300 Evolutionary history of tissue bending

Newly discovered unicellular eukaryote shows hallmark feature of animal morphogenesis *By P. Tomancak*
RESEARCH ARTICLE p. 326

302 Mobilizing unconventional T cells

Modulating unconventional antigen presentation could treat infections and cancer *By T. H. M. Ottenhoff and S. A. Joosten*

304 Photocatalytic deracemization fixes the mix

Photocatalysis converts racemic mixtures of *N*-aryl cyclic ureas into single enantiomers *By A. E. Wendlandt*
REPORT p. 364

305 Photosystem II, poised for O₂ formation

A consensus is emerging on the structure of the oxygen-evolving complex *By R. D. Britt and D. A. Marchiori*
RESEARCH ARTICLE p. 334

306 Spikes in the sleeping brain

Memory is replayed and consolidated under low background noise during deep sleep *By Y. Ikegaya and N. Matsumoto*
REPORT p. 377

308 Rapid reorganization of global biodiversity

Marine systems outpace terrestrial habitats in biodiversity erosion *By B. K. Eriksson and H. Hillebrand*
RESEARCH ARTICLE p. 339

POLICY FORUM

310 Technological challenges and milestones for writing genomes

Synthetic genomics requires improved technologies *By N. Ostrov et al.*

BOOKS ET AL.

313 The work of words in the Anthropocene

Two books—one fictional, one dictionary—probe how we conceive of our changing world *By D. Dixon*

314 The stories that make us spend (and save)

A Nobel laureate reveals how narratives shape the economy *By S. Jaffe*

LETTERS

315 Forest restoration: Overlooked constraints

By E. Luedeling et al.

316 Forest restoration: Expanding agriculture

By R. Delzeit et al.

316 Forest restoration: Transformative trees

By D. Sheil et al.

317 Response

By J.-F. Bastin et al.

318 Technical Comment abstracts

RESEARCH

IN BRIEF

320 From *Science* and other journals

REVIEW

323 Signaling

Sphingosine 1-phosphate: Lipid signaling in pathology and therapy *A. Cartier and T. Hla*

REVIEW SUMMARY; FOR FULL TEXT:
DX.DOI.ORG/10.1126/SCIENCE.AAR5551

RESEARCH ARTICLES

324 Human evolution

Adaptive archaic introgression of copy number variants and the discovery of previously unknown human genes *P. Hsieh et al.*

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:
DX.DOI.ORG/10.1126/SCIENCE.AAX2083

325 Neurodevelopment

Coordination between stochastic and deterministic specification in the *Drosophila* visual system
M. Courgeon and C. Desplan

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:
DX.DOI.ORG/10.1126/SCIENCE.AAY6727

326 Cell biology

Light-regulated collective contractility in a multicellular choanoflagellate *T. Brunet et al.*

PERSPECTIVE p. 300



334 Water oxidation

An oxyl/oxo mechanism for oxygen-oxygen coupling in PSII revealed by an x-ray free-electron laser *M. Suga et al.*

PERSPECTIVE p. 305

339 Biodiversity

The geography of biodiversity change in marine and terrestrial assemblages
S. A. Blowes et al.

PERSPECTIVE p. 308

346 Earthquakes

Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence
Z. E. Ross et al.

351 Gene expression

Genetic regulatory variation in populations informs transcriptome analysis in rare disease
P. Mohammadi et al.

REPORTS

356 Exoplanets

Oxygen fugacities of extrasolar rocks: Evidence for an Earth-like geochemistry of exoplanets *A. E. Doyle et al.*

360 3D printing

Rapid, large-volume, thermally controlled 3D printing using a mobile liquid interface *D. A. Walker et al.*

364 Organic chemistry

Light-driven deracemization enabled by excited-state electron transfer *N. Y. Shin et al.*

PERSPECTIVE p. 304

Quantum computing

369 Deterministic generation of a two-dimensional cluster state *M. V. Larsen et al.*

373 Generation of time-domain-multiplexed two-dimensional cluster state *W. Asavanant et al.*

377 Neuroscience

Isolated cortical computations during delta waves support memory consolidation
R. Todorova and M. Zugaro

PERSPECTIVE p. 306

382 Turbulence

Synthetic dissipation and cascade fluxes in a turbulent quantum gas
N. Navon et al.

DEPARTMENTS

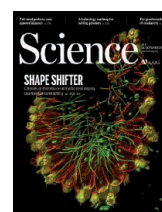
281 Editorial

Scientists and politics?
By Ian L. Boyd

390 Working Life

Escaping 'The Waiting Place'
By Nathan Pickle

ON THE COVER



A colony of the protozoan species *Choanoflexa flexa*. As a choanoflagellate, *C. flexa* is one of the closest living relatives of animals. This species forms cup-shaped colonies that quickly invert

their curvature in response to light-to-dark transitions by undergoing collective cell contractions. This behavior may offer insight into the evolutionary origin of tissue bending in animals. See pages 300 and 326. *Image: Thibaut Brunet (University of California Berkeley/Howard Hughes Medical Institute)*

New Products..... 386
Science Careers387

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2019 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): \$165 (\$74 allocated to subscription). Domestic institutional subscription (51 issues): \$1971; Foreign postage extra: Mexico, Caribbean (surface mail) \$55; other countries (air assist delivery): \$98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request. GST #R125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.** **Change of address:** Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090-6178. **Single-copy sales:** \$15 each plus shipping and handling; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Scientists and politics?

The lower legislative chamber of the United Kingdom's Parliament has 650 members, but only one (0.15%) has a science Ph.D. This seems like a surprisingly small number in a mature democracy. About 0.8% of people in the United Kingdom have a science Ph.D., so it appears that science is seriously underrepresented. I suspect it is the same the world over. Why is this, is it right, and what are the consequences?

There is, fortunately, an increasing focus on making governments representative of the diversity of the population they serve in terms of gender, race, and sexual orientation. But diversity also needs to embrace different intellectual approaches. The structured thinking and disciplined methodologies of science add to diversity, but these are aspects that can challenge vested interests. The blunt, socially insensitive, scientist speaking truth to power is certainly a caricature, but it is sufficiently real to warrant careful management by governments. There is also often suspicion that scientists operate their own agendas.

Vested interests do not want their political, social, and financial currency debased by being confronted by the real world. Governments have, therefore, tended to put scientists in a metaphorical box and to only open the lid when they are needed, thus reducing science to a technical service function, supplying support, advice, and economic goods. These boxes can take many forms, from the containment of advisers with a set of rules to forums created for science to play within, like the Intergovernmental Panel on Climate Change (IPCC). Until recently, I was chief scientific adviser in the U.K. government on food and environment, a role that only slightly corrected for poor integration of science within government.

It is this systematic exclusion that leads to the underrepresentation of scientists in politics and government. Scientific culture is dominated by human social culture and is forced to play by its rules. As C. P. Snow said 60 years ago, "This polarisation is sheer loss to us all." Politics would look very different if it were more evenly

balanced between these cultures. Aristotle saw politics as legislative science or learning by experience. Politics was itself a scientific inquiry, thus reversing the current maxim that science lies outside politics. In Aristotle's world, scientists would be those who designed policies to solve tricky problems.

The modern idiom of scientists as the custodians and discoverers of knowledge is much too restrictive. When contested issues arise, society needs scientists to be dominant voices acting as authoritative, impartial, and trusted arbiters who can explain where uncertainties lie in evidence and the likely consequences of alternative futures.

Unfortunately, with the possible exception of climate change (helped by the IPCC), scientists are left to struggle to form a politically influential caucus working within its own rules. The alternative is capture by the language and ways of the other culture. Under this scenario, science becomes yet another money-grabbing vested interest. As a result, many of those who would most benefit from listening put scientists in an opposing political box and close the lid. Advocacy is the surest and most rapid way to achieve such an effect. It can compromise the impartiality of scientific evidence and do more harm than good. When science becomes captured by

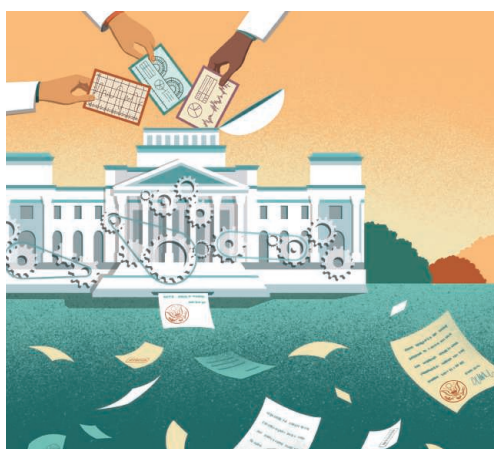
normal politics, its value drains away.

Unless the scientific community grips this problem of systemic disempowerment within government and politics, science will continue to be manipulated within a political game dominated by vested interests. The underrepresentation of science seems likely to continue unless the process of government itself becomes more diverse. Scientists themselves can help by demonstrating greater thought leadership, showing more enthusiasm to become part of government in all its forms, and valuing the contributions of colleagues who get involved. They need to break out of their synthetic box and politely refuse to have the lid closed on them.

—Ian L. Boyd



Ian L. Boyd is a professor of biology at the University of St Andrews, Fife, Scotland, UK, and chairman of the UK Research Integrity Office. He is the former chief scientific adviser in the United Kingdom on food and environment. ilb@st-andrews.ac.uk



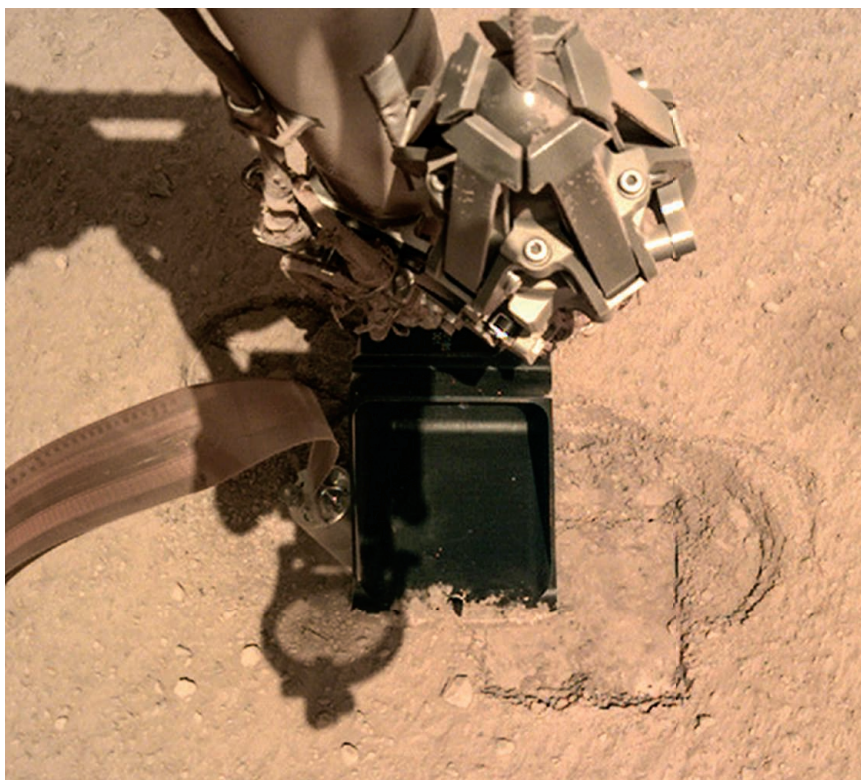
"...systematic exclusion...leads to the underrepresentation of scientists in politics and government."

“It'll have to be bigger, brighter and pay better.”

Statistician Adrian Smith, author of a forthcoming report recommending the United Kingdom create an alternative to the European Research Council if U.K. scientists cannot receive its funding post-Brexit.

IN BRIEF

Edited by **Jeffrey Brainard**



The InSight lander's robot arm, shown hovering over its heat probe, is helping it penetrate the surface.

PLANETARY SCIENCE

Mars lander resumes hammering probe

NASA last week reported progress in getting the stuck heat probe on its InSight lander to resume hammering itself below the martian surface. During the previous 9 months, the probe, designed to measure heat flow from Mars's interior, had penetrated only 35 centimeters instead of the planned 5 meters. The lumpy soil surrounding the hole does not cave into it as the probe tip hammers, preventing the probe from gaining the traction necessary for penetration. After rescue attempts failed this summer, NASA and scientists from the German Aerospace Center (DLR), which designed the probe, finally decided to use InSight's robotic arm to pin the heat probe, called a mole, against the side of its hole, increasing friction. Hammering resumed last week, and the mole had already penetrated 3 centimeters more when *Science* went to press, DLR said.

Econ experimenters win Nobel

PRIZES | This year's Nobel Memorial Prize in Economic Sciences honors three pioneers of the use of randomized controlled trials to determine how best to ameliorate global poverty. Michael Kremer of Harvard University and Abhijit Banerjee and Esther Duflo of the Massachusetts Institute of Technology in Cambridge have often worked together to test antipoverty interventions in education, health, agriculture, and access to credit, mainly in sub-Saharan Africa and India. The experimental approach in economics is labor intensive and faces questions about whether effects observed in such trials hold over the long run. Duflo, 46, is just the second woman to win the economics Nobel and the only woman among the 14 Nobel laureates this year.

Chemical attack paper vetoed

WARFARE | In an about-face, the journal *Science & Global Security* decided last week not to publish a controversial article that casts doubt on the Syrian government's responsibility for a 2017 chemical attack that killed more than 80 people. Scientists had accused one of the authors, professor emeritus Ted Postol of the Massachusetts Institute of Technology in Cambridge, of pushing conspiracy theories. The journal's editors wrote on its website that they cannot “rectify” problems with the manuscript or provide it “an independent, fair, effective, and conclusive blind peer review.” Postol called the decision “totally wrong” and said he would resign from the journal's editorial board. (He says he has not been involved in the journal's deliberations about the paper.) An email from the editors to the authors, which Postol shared with *Science*, says the paper was sent to two reviewers who were divided, but does not provide details.

Theft of biblical texts alleged

ANTIQUITIES | Dirk Obbink, a professor at the University of Oxford in the United Kingdom, was accused this week of stealing 11 ancient Egyptian papyrus fragments from a scholarly collection he directed and selling them to Hobby Lobby Stores, the U.S. craft chain run by the evangelical

ARCHAEOLOGY

Ancient Maya staircase preserved by replica

Researchers this week said they will soon protect an intricately carved, 1000-year-old Maya staircase by capping and covering over its unique hieroglyphic panels with replicas. The staircase is in a palace patio in the ancient city of Palenque in southern Mexico. In the 1890s, British explorer Alfred Maudslay took plaster molds of the staircase's glyphs and made other records of Palenque and other Maya cities. Rain and sun have since taken a toll on the staircase. Over the past 3 years, the British Museum in London worked with Google Arts & Culture to digitally scan Maudslay's archive, which the museum holds. The scans (shown in a digital model, right) were then used to mill exact replicas of the glyphs from limestone, the same material used for the original.

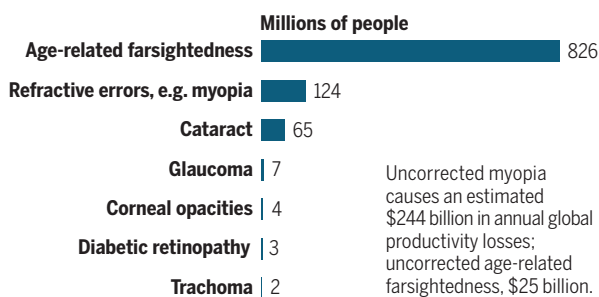


Green family. The fragments, all inscribed with biblical texts, ended up at the Museum of the Bible in Washington, D.C., which the Greens built in part to display their collection of such artifacts. That museum and the Egypt Exploration Society (EES), a London nonprofit that originally owned the fragments, found that 13 fragments had mysteriously gone missing from EES; 11 can be traced to sales by Obbink to Hobby Lobby, EES reported on 14 October. The museum says it will return all 13 fragments to EES. Efforts to reach Obbink for comment were unsuccessful.

Vision loss goes unaddressed

PUBLIC HEALTH | At least 1 billion people live with a moderate to severe vision impairment, such as glaucoma or farsightedness, that is uncorrected or could have been prevented, says a report last week by the World Health Organization (WHO). The burden is disproportionately high in

▲ Avoidable impairment



women and other specific populations, including people in rural areas and low- and middle-income countries; poor residents of sub-Saharan Africa and South Asia have rates of blindness eight times higher than people in high-income countries, WHO says. The number of unaddressed vision problems will likely increase significantly in coming years as the world's population ages and more people move to cities, the report says; urban living is associated with increased rates of nearsightedness. More than \$20 billion would be needed to correct or prevent the unaddressed vision impairments. The report recommends including vision care in primary health care settings and national health plans.

India tightens China exchanges

INTERNATIONAL AFFAIRS | India moved last week to scrutinize its collaborative research agreements with institutions in China. Under a new policy from the University Grants Commission, which regulates more than 900 institutions of higher education, India's ministries of Home Affairs and External Affairs need to approve such deals in the future. Scientists fear the policy, triggered by tensions over borders and growing mistrust between the countries, will further squeeze

what is already a trickle of such exchanges: From 1995 to 2018, only 88 Indian scientists visited China and 73 Chinese researchers came to India under a program run by the countries' national science academies. (A small number of additional exchanges have occurred under other auspices.)

Congo measles cases lead globe

INFECTIOUS DISEASE | The Democratic Republic of the Congo (DRC), already struggling with an outbreak of the Ebola virus, is also contending with the world's largest current surge of measles, with more than 203,000 cases so far this year, triple its 2018 total, the United Nations Children's Fund (UNICEF) announced on 9 October. Children under age 5 are bearing the brunt of the epidemic, accounting for 74% of infections and nearly 90% of the country's 4096 measles deaths this year. (By comparison, 2143 people have died from Ebola since the current outbreak emerged in the summer of 2018.) In 2018, measles vaccination coverage in the DRC was just 57%; about 95% is required to prevent measles outbreaks. Weak health care infrastructure and community mistrust of vaccines and vaccinators have contributed to the poor coverage, UNICEF said.

U.S. STDs break record

PUBLIC HEALTH | Combined total U.S. cases of the three most commonly reported sexually transmitted diseases (STDs)

THREE QS

A bias in Nobels to women

This year, only one woman won a Nobel Prize in a scientific field, economics. Since the awards were first given in 1901, women have taken home just 22 Nobels, about 3% of the total. Is that because women were simply underrepresented in research fields years or decades before the prizes were given? According to physicist Liselotte Jauffred of the University of Copenhagen, the answer is no. With two colleagues, she published an analysis in May in *Palgrave Communications* showing that bias, not underrepresentation, almost certainly accounts for the scarcity of women Nobel winners.

Q: How did you do your analysis?

A: We assumed that among all tenured faculty members, everybody would have the same chance to enter this category of really esteemed researchers. We sampled a pool of [U.S. National Science Foundation data on] faculty members and said that if there is a 10% chance of the faculty members that are women, then we would think that the success rate of women to get a Nobel Prize would also be 10%. Then we looked at the success rate of women in getting the Nobel Prize and see that it is much, much lower than what the gender ratio suggests.

Q: Can your study say anything about the sources of bias?

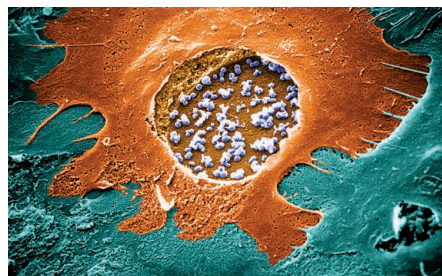
A: Our study is purely a statistical study that says that there is a bias. [But] I don't actually think it's so much the Nobel Committee that is implementing this bias. I think it's something that happens in multiple earlier steps.

Q: The Royal Swedish Academy of Sciences said it has increased the number of female nominees for the prizes. Are there signs of progress?

A: We cannot say anything about trends based on a single year. We did make a very small Monte Carlo simulation, though. Every year, the chance that the winners will be all men in chemistry, economics, and physics is over 80%, and medicine is around 60%. ... In the last 20 years, there have been about the same number of women Nobel laureates as in the first 100 years. It's progressing in the right direction, but not fast enough.

S SCIENCEMAG.ORG/NEWS
Read more news from *Science* online.

climbed to an all-time high of nearly 2.5 million in 2018, the Centers for Disease Control and Prevention (CDC) announced on 8 October. Cases of chlamydia grew by 3% from the 2017 level, to more than 1.7 million, the most ever. Gonorrhea cases increased 5%, to 583,405, the highest since 1991. And cases of early- and middle-stage syphilis, the most infectious stages of the disease, grew 14% to 35,063, also the most since 1991. Ninety-four newborns died from congenital syphilis, which is passed from mothers to babies during pregnancy, up 22%. All three diseases are readily treatable with antibiotics. CDC said the increases were caused in part by cuts to STD screening and treatment programs and decreased condom use.



Untreated, the chlamydia bacterium can cause infertility in women.

A bid to renew stem cell fund

RESEARCH BUDGETS | As California's \$3 billion stem cell agency prepares to dole out the last research awards from its dwindling funds, the investment banker behind the 2004 ballot initiative that created it has formally launched a new effort to keep it alive. As anticipated, Robert Klein, chairman of the advocacy group Americans for Cures in Palo Alto, California, submitted a new ballot measure last week that would devote another \$5.5 billion from bond sales to the California Institute for Regenerative Medicine (CIRM), which stopped accepting nearly all new award applications at the end of June. The measure includes proposals to increase public access to potential therapies generated by CIRM-funded research, including the creation of new centers to host clinical trials across California. If the petition can win the required 623,212 signatures, it will appear on the ballot in November 2020.

Ex-EPA air panel weighs in

POLLUTION | The U.S. Environmental Protection Agency (EPA) should tighten its limits on particulate air pollution, an unusual group of 20 former agency science advisers has concluded. The experts had

served on an official EPA advisory panel that the agency disbanded last year, saying it was no longer needed. Critics said the dismissal was aimed at easing Trump administration efforts to weaken particulate limits, and some former panelists decided to independently continue their work with help from the Union of Concerned Scientists, an advocacy group in Cambridge, Massachusetts. After meeting on 10 and 11 October near Washington, D.C., the group expects to recommend that EPA cut its main fine-particulate limit from 12 micrograms per cubic meter ($\mu\text{g}/\text{m}^3$) of air to between 8 and 10 $\mu\text{g}/\text{m}^3$, *E&E News* reported. EPA can ignore the advice, but the former advisers hope their work will bolster legal efforts to force the agency to impose tighter limits.

Psychology studies criticized

SCHOLARSHIP | An investigation by King's College London has deemed 26 papers by the late Hans Eysenck, a prominent U.K. psychologist, "unsafe," and the university has recommended they be retracted, *The Guardian* reports. Eysenck's controversial theories, many formed in collaboration with German researcher Ronald Grossarth-Maticek in the mid-20th century, included his belief that people with certain personality types are more prone to cancer and heart disease, and that smoking might not actually be carcinogenic. Other experts criticized his publications, which were nevertheless widely cited—Eysenck ranked third in the social sciences citation index behind Sigmund Freud and Karl Marx. The investigation was prompted by a paper published in February by psychiatrist Anthony Pelosi of Priory Hospital in Glasgow, U.K., criticizing Eysenck's work.

Drexel settles bar bill case

OVERSIGHT | An investigation by Drexel University found that its former head of electrical and computer engineering spent \$189,062 of federal grant money at strip clubs and sports bars and for other non-work-related expenses over 10 years. Drexel, in Philadelphia, Pennsylvania, told the government that Chikaodinaka Nwankpa charged bills at Club Risque and other establishments from 2007 to 2017 to eight federal grants for energy and naval technology research, the U.S. Attorney's Office for the Eastern District of Pennsylvania announced 7 October. In a settlement with the government, Drexel agreed to pay back the money; Nwankpa resigned his position, repaid the university \$53,328, and was barred from federal government contracting for 6 months, the U.S. Attorney's Office said.



IN DEPTH

The Biosphere 2 rainforest greenhouse, built in Arizona in the late 1980s, contains 90 plant species.

ECOLOGY

Drought test begins in Biosphere 2 rainforest

Tracing carbon in closed ecosystem could reveal tree and soil response to water stress

By Erik Stokstad

Earlier this month, the doors to the tropical rainforest, enclosed under a ziggurat of glass, were sealed shut. Christiane Werner turned a valve to release about \$12,000 worth of carbon dioxide (CO₂) spiked with carbon-13, an isotope that is normally scarce in the atmosphere. The luxuriant plants inside Biosphere 2, a 30-year-old set of greenhouses and artificial ecosystems in the Arizona desert, soaked up the isotopic tracer, enabling investigators to follow the flows of carbon through the healthy forest. Werner, an ecosystem physiologist at the University of Freiburg in Germany, and her team gathered these baseline data for the harsh test to come: the largest forest drought experiment ever monitored with isotopes. “It will be amazing to see the results,” says Tamir Klein, a plant ecophysiologicalist at the Weizmann Institute of Science in Rehovot, Israel, who is not involved.

On 7 October, the researchers shut off the sprinklers that irrigate the rainforest, beginning a 6-week drought. Next month, they will inject another pulse of isotopically enriched CO₂ into the densely instrumented ecosystem, and apply other tracers. A forest’s consumption of CO₂ slows during drought, but scientists haven’t pinned down how thirsty rainforest plants—especially large trees—use and release their stored carbon. The answers are important for the

global climate cycle, Klein says. Droughts, expected to become more severe as the climate warms, could turn tropical forests from sinks of greenhouse gases into sources that accelerate climate change.

Field experiments in the Amazon, in which plastic panels intercept rain to keep large swaths of forest dry, have sketched out how drought kills trees of different sizes (*Science*, 15 April 2005, p. 346). Smaller studies targeting individual plants with isotopic tracers have revealed some of the impacts on plant function. But the Biosphere 2 experiment will do both by applying tracers across an entire forest. “We have an ecosystem in a lab,” Werner says.

The \$150 million Biosphere 2 was built in the late 1980s as a kind of spaceship on Earth, in which humans would attempt to survive inside a sealed ecosystem. That

mission flopped, but the University of Arizona now operates the facility for research, education, and tourism. It has hosted large ecology studies and an ongoing \$3 million experiment in landscape evolution (*Science*, 14 December 2012, p. 1417). Biosphere 2’s original funder, financier Edward Bass, helped support that earlier work, but much of the new experiment is funded by part of a €1.9 million grant Werner won from the European Research Council. About 50 researchers from 13 institutions are contributing equipment and expertise.

The focus is Biosphere 2’s tropical forest, which includes some 90 plant species across an area the size of seven tennis courts. All summer, the team prepared by building canopy platforms where they could enclose dozens of leaves and stems in small chambers to capture their emissions. They drilled into tree trunks to insert probes, and dug observation pits to measure emissions from soil and roots. Four kilometers of tubing carry gases from the probes to a room full of instruments. “The scale of measurements on this drought is completely unparalleled,” says co-leader Laura Meredith, a biogeochemist at the University of Arizona in Tucson and director of rainforest research at Biosphere 2.

By tracking the carbon-13, the researchers will learn how quickly carbon is taken up during photosynthesis and then moves through the forest. They will compare those rates before and during the drought across six tree species that differ in their



Gases from leaves provide clues to whether thirsty rainforests are carbon sources or sinks.

drought resistance. And they will learn how the trees apportion stored carbon in their leaves, trunks, and roots. It's a "huge black box," and crucial for predicting how plants respond to stresses like drought, says plant physiologist William Anderegg of the University of Utah in Salt Lake City.

Another set of tracers will show in finer detail how particular metabolic pathways use carbon. During the past month, the researchers have supplied a solution of isotopically enriched pyruvate, a chemical building block used in many biological processes, to leaves, roots, and clumps of soil. One type of pyruvate tracer reveals how much carbon is given off during daytime respiration—a key part of the carbon cycle that needs to be better quantified, Werner says.

Another pyruvate tracer, taken up into a different pathway, shows how much carbon the plants and soil microbes use to synthesize volatile organic compounds (VOCs). When plants are stressed, these chemicals make up a considerable fraction of their carbon emissions. They can warm the atmosphere or turn into aerosols that cool it, but their overall climatic effect is unknown. Plants use VOCs for many purposes, including as a homing signal for a vast web of soil fungi that provide water and nutrients to roots during drought. The researchers hope to quantify rates and amounts of VOCs exchanged between the microbes and plants and whether they change during drought.

At the end of the drought, the researchers will perform one last tracer experiment, irrigating the deep soil with water enriched in an isotope of hydrogen. They expect large trees to take up most of the water, and they hope to learn whether their deep root systems will leak some of the water into the shallow soil, helping smaller plants recover.

Finally, the sprinklers will turn on and return the ecosystem to normal. When parched soil and fallen leaves are rewetted, microbes go into metabolic overdrive and churn out CO₂ and VOCs. Meredith and her colleagues will measure emissions and link them to patterns in microbial genes.

Ultimately, results from the drought test will improve the way global climate models account for vegetation. "You need these experiments to unlock the physiology and add it into the models," Anderegg says. "It gets us much more mechanistic and rigorous projections of how tropical trees and forests might respond to climate change."

After the experiment wraps up, tourists will be let into the rainforest again. But the canopy platforms will remain for future research, and some of the carbon tracers will also stick around. "We can look for the signal for years to come," Meredith says. ■

SCIENTIFIC COMMUNITY

NIH 'high risk, high reward' awardees skew male—again

Agency struggles to attract female applicants, keep review process gender neutral

By Meredith Wadman

When the U.S. National Institutes of Health (NIH) announces each year's winners of plum "high risk, high reward" (HRHR) awards, which go to about 100 scientists doing outside-the-box research, the paucity of female winners regularly raises eyebrows. Despite NIH's efforts to improve the picture, this year's winners, announced on 1 October, were no exception.

Although women won three of the four types of HRHR awards in numbers that met or exceeded their representation in the applicant pool, that representation was meager: For all HRHR awards, 18% to 38% of applicants were women, although women have earned more than 50% of U.S. Ph.D.s in biological sciences since 2008. What's more, for the Early Independence Awards (EIAs)—which can boost women's careers at a crucial time by allowing new Ph.D.s to skip a postdoc and start an independent lab immediately—women constituted 38% of applicants, but only 25% of awardees.

The EIA sample size is small: Among the 13 new awardees, three were women. Still, women have been underrepresented among EIA winners in eight of the 9 years since the award's launch. "In aggregate over all the years, there has been a significant bias," says Kristin Knouse, a 2018 EIA winner who is a cell biologist at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts. "There needs to be a systematic examination of where this awardee bias is arising."

But Olivia Corradin, also at the Whitehead Institute, who in 2017 won a similar early-investigator award for innovative research from NIH, cautions against reading too much into the data. "You could pull the same analysis for the [HRHR] Pioneer Award and you might conclude that they are overrepresenting women."

Women have won Pioneer Awards in numbers that met or exceeded their representa-

tion in the applicant pool in seven of the past 9 years. This year, 18% of applicants—and 45% of awardees—were women.

The low numbers of women applying for all four awards are the main concern, says James Anderson, who directs NIH's Division of Program Coordination, Planning, and Strategic Initiatives, which supports the HRHR awards. "We need to focus on encouraging more women to apply," he says. "We take this pretty seriously. We gotta change these numbers." (For the two other types of HRHR awards this year, the Transformative Research and New Innovator awards, women were 20% and 30% of applicants, respectively, and won 24% and 33% of awards, respectively.)

In 2018, concerned about disparities among HRHR applicants, NIH Director Francis Collins convened a working group to review the program. Their final report, issued this past June, urged vigorous outreach to institutions and populations that historically have not applied for HRHR awards in large numbers. The report also flagged a bias in favor of awards to scientists at high-profile institutions, and for proposals to study mechanistic and subcellular topics over, for instance, behavioral and clinical research.

In April, NIH tweaked its call for 2020 applications. "We strengthened our inclusivity language," says HRHR Program Leader Ravi Basavappa, to make clear that the agency welcomes research from the whole spectrum of institutions, all sorts of applicants, and any topics within its mission.

Both the stubbornly low numbers of women applying for HRHR awards and their poor record at winning EIAs show "just how difficult change is," says Molly Carnes, a working group member who is an expert on implicit bias and a professor of women's health at the University of Wisconsin in Madison. "We live gendered lives. When you evaluate men and women's science, their science is filtered through the fact that they are men and women. That's how the human mind works." ■

"We take this pretty seriously. We gotta change these numbers."

James Anderson,
National Institutes
of Health



Sauropods, the largest terrestrial animals ever to have lived, may have relied on foods more nutritious than ferns and conifers to grow to massive size.

PALEONTOLOGY

Sauropods get a new diet and a new look

The behemoths may have chomped nutritious horsetails—with the help of sturdy beaks

By **John Pickrell**, in Brisbane, Australia

How did sauropod dinosaurs, the biggest creatures ever to have thundered across Earth, bulk up to the weight of more than 10 African bull elephants on a spartan diet of prehistoric greens? Many herbivores today grow fat on energy-rich grasses, but these and other nutritious flowering plants didn't become common until near the end of the dinosaurs' reign. Now, researchers think they have glimpsed the answer: a surprisingly nutrient-rich plant that could have been a mainstay of these dinosaurs' diets, and turtlelike beaks that buttressed sauropods' peglike teeth as they relentlessly stripped foliage from plants.

"We are seeing that they were able to acquire large volumes of food rapidly, and possibly nutrient-rich food," says Stephen Poropat, a paleontologist at Swinburne University in Melbourne, Australia.

Researchers at the University of Bonn in Germany presented the findings last week at the annual meeting of the Society of Vertebrate Paleontology here. Their previous work helped show that sauropods were eating machines that gulped down vegetation without chewing. They swung their long necks over vast areas, like prehistoric lawn mowers, while saving energy by keeping their bodies in one spot. The new studies add detail by exploring the beasts' diet and jaw structure.

One study, by Bonn's Carole Gee and her colleagues, identified what may have been the superfood of the dinosaur era. Her team zeroed in on the nutritional content of low-growing, spore-bearing horsetails, or *Equisetum*,

which were widespread during the Jurassic period and still grow today.

Horsetails appeared to be poor fodder in previous tests, which simply burned the plants to measure carbon content, Gee says. Instead, her team adapted the Hohenheim gas test, a method for assessing the quality of fodder for farm animals. They fermented modern horsetails for 3 days to simulate the journey through a sauropod's gut and measured the volume of gas produced—an indicator of energy content. The researchers were astounded to find that horsetails released more energy than any other plant group, including 16 modern grasses. *Equisetum* is rich in protein, they say, and far more nutritious than the ferns, cycads, and conifers common in the dinosaur era. Gee argues that horsetails by rivers and lakes would have offered sauropods, especially young ones, "a plentiful, accessible, and extremely nutritious food."

Even eating superfoods, sauropods must have vacuumed up as much as 1 ton or more of plant matter per day. To understand the mechanics, another Bonn team analyzed the beasts' jaws.

For decades, paleontologists have dug up puzzling fossils: rows of isolated sauropod teeth, still neatly arranged as they would have been in the mouth, with not a scrap of bone encasing them. "There must have been something holding them in place," Bonn's Kayleigh Wiersma says. "Otherwise they would have been scattered all around the dig site."

In some skulls that retain teeth, they seem to be almost falling out of the sockets. "You can't have teeth exposed to that degree," says

Steve Salisbury, a paleontologist at the University of Queensland here. "It seems likely there would have been some tissue that enclosed the base at least."

Wiersma and research leader Martin Sander studied seven sets of tooth rows, with up to 40 teeth each, from sauropods including *Diplodocus*, *Brachiosaurus*, and *Apatosaurus*. Wiersma said at the meeting that paleontologists may have had the faces of these iconic dinosaurs all wrong: Instead of the lizardlike lips shown in artists' concepts, the behemoths likely sported beaks as well as teeth, unlike any living animal.

The researchers found that sauropod teeth typically show surface wear only about halfway down to the jaw. That suggests the teeth were once deeply embedded in a supporting structure. The researchers also found tiny pits in the surface of the jaw, perhaps indicating blood vessels to nourish beaks. Similar pits and foramina are seen in other dinosaur bones near inferred structures, like the sheaths covering horns, thought to have been made of keratin, which forms our nails and birds' beaks and feathers.

Poropat cautions that expanded gum tissue, rather than a beak, might have held the teeth in place. But bony scaffolding that supported beaks is obvious in other dinosaurs, including *Triceratops*, *Stegosaurus*, and duck-billed hadrosaurs. Nevertheless, "I don't think we would have expected sauropods to have beaks," says Darren Naish, a paleontologist at the University of Southampton in the United Kingdom. It's a "whole new look." ■

John Pickrell is a journalist in Sydney, Australia.

CHEMISTRY

Lithium-ion battery development takes Nobel

Three chemists honored for creating powerful, light, rechargeable cells

By Robert F. Service

You probably have the work of Nobel Prize winners in your pocket. This year's Nobel Prize in Chemistry was awarded last week to the pioneers of the lithium-ion battery, an invention that has become ubiquitous in the wireless electronics that permeate modern life: your phone, tablet, laptop, and perhaps even your car. Lighter and more compact than the rechargeable batteries that preceded them, lithium-ion batteries are now moving beyond gadgets to power homes, airplanes, and even the electric grid that feeds power to entire cities.

Three chemists will split the \$900,000 prize: Stanley Whittingham at the State University of New York in Binghamton, John Goodenough at the University of Texas in Austin, and Akira Yoshino at Asahi Kasei Corporation in Tokyo.

"I am thrilled," says Yang Shao-Horn, an electrochemist at the Massachusetts Institute of Technology in Cambridge. "Lithium-ion batteries have made a tremendous impact on our society," adds Kristin Persson, a materials chemist at the University of California, Berkeley: "All three [awardees] deserved it. They provided critically important components to make lithium-ion batteries work."

Virtually all batteries have three essential components: two electrodes—an anode and a cathode—separated by an electrolyte. In today's lithium-ion cells, the electrolyte is typically a liquid that allows lithium ions to move back and forth between the electrodes. When the battery is providing electricity, lithium atoms at the anode give up electrons that flow through an external wire. The positively charged lithium ions left behind migrate through the electrolyte to the cathode, where they nestle between the cathode's layers of metal oxide materials. Applying an electric voltage to charge the battery reverses the flow, pushing the lithium ions out of the metal oxide, through the electrolyte, and back to the anode, where they retrieve electrons.

In the 1970s, when Whittingham was working at Exxon, all rechargeable batteries used other, less efficient electrode materials:

lead, nickel, cadmium. Whittingham tried something new: pairing a cathode made of a layered material called tantalum disulfide (TaS₂) with an anode made from lithium metal, which readily gives up electrons and forms ions that can wedge into the TaS₂ layers. The resulting battery could store lots of energy. But Exxon couldn't commercialize it because repeated charging deposited thin whiskers of lithium, known as dendrites, on the anode, causing some batteries to short out and catch fire.

Goodenough, then at the University of Oxford in the United Kingdom, picked up the mantle. He realized the cathode could soak up more electrons if it was made of a metal oxide instead of a metal sulfide. Some of those proved unstable. But he also found

"I just sort of sniffed out the direction the trends were moving," Yoshino says. By 1991, Asahi Kasei and Sony had released the first commercial lithium-ion battery, which powered portable music players, among other devices.

Those initial cells could store twice as much energy as the previous best rechargeable batteries. Further improvements have boosted the energy storage capacity of lithium-ion cells another threefold, notes George Crabtree, director of the Joint Center for Energy Storage Research at Argonne National Laboratory in Lemont, Illinois. "The progress has been absolutely stunning," he says.

It's likely not over. "There's still plenty of room for improvement," Persson says.

For starters, battery researchers are working feverishly to do away with cobalt, an expensive metal often mined by child laborers in the Democratic Republic of the Congo, which supplies some 70% of the world's total. Current lithium-ion batteries reduce the need for cobalt by mixing in manganese and nickel, but stabilizing the layered cathode still requires some cobalt. "We have to move away from that," Persson says.

Another major research thrust is replacing liquid electrolytes with polymers and other solid materials. If perfected, solid electrolytes could enable researchers to return to using solid lithium anodes without fear of creating dendrites. That could lead to bat-

teries able to store enough energy to propel electric vehicles as far as gas-powered cars without recharging, power short-hop electric planes, and even store backup electricity from wind and solar plants to light up cities at night and when the wind is still.

Goodenough, 97, is the oldest person ever to receive a Nobel Prize. He was in London to accept another award when he got word. Asked during a press conference what he hopes for on his return to his lab in Texas, he said, "I hope they will still keep me employed." Whittingham, 77, has no plans to cut back his research either. "I'm going to follow John Goodenough's example," he says. "I'm still full steam ahead." The same is certainly true for their creation. ■



The 1970s oil crisis spurred interest in rechargeable batteries; today's electric car batteries use technology launched from the chemistry Nobelists' ideas.

that adding cobalt to the cathode helped stabilize its layers. In 1980, he described a rechargeable battery that could produce 4 volts, making it nearly twice as powerful as Whittingham's.

Still, dendrite formation on lithium metal anodes remained a major problem. Researchers had shown they could make anodes out of graphite instead, because lithium atoms could wedge themselves between the graphite's carbon layers. But the graphite invariably broke down in the battery's electrolyte. In the mid-1980s, Yoshino found that if he used an electrolyte containing ethylene carbonate, a thin insulating layer formed on the graphite surface, protecting it from degradation.



NEUROSCIENCE

Rival theories face off over brain's source of consciousness

\$20 million project puts competing ideas to the test

By Sara Reardon

Brain scientists can watch neurons fire and communicate. They can map how brain regions light up during sensation, decision-making, and speech. What they can't explain is how all this activity gives rise to consciousness. Theories abound, but their advocates often talk past each other and interpret the same set of data differently. "Theories are very flexible," says Christof Koch, president of the Allen Institute for Brain Science in Seattle, Washington. "Like vampires, they're very difficult to slay."

Now, the Templeton World Charity Foundation (TWCF), a nonprofit best known for funding research at the intersection of science and religion, hopes to narrow the debate with experiments that directly pit theories of consciousness against each other. The first phase of the \$20 million project, launched this week at the Society for Neuroscience meeting in Chicago, Illinois, will compare two theories of consciousness by scanning the brains of participants during cleverly designed tests. Proponents of each theory have agreed to admit it is flawed if the outcomes go against them.

Head-to-head contests are rare in basic science. "It's a really outlandish project," says principal investigator Lucia Melloni, a neuroscientist at the Max Planck Institute for Empirical Aesthetics in Frankfurt, Germany. But understanding consciousness has become

increasingly important for researchers seeking to communicate with locked-in patients, determine whether artificial intelligence systems can become conscious, or explore whether animals experience consciousness the way humans do. To winnow the theories, TWCF took inspiration from a 1919 experiment in which physicist Arthur Eddington pitted Albert Einstein's theory of general relativity against Isaac Newton's gravitational theory. Eddington measured how the Sun's gravity caused light from nearby stars to shift during a solar eclipse—and Einstein won.

Consciousness, however, is much less easily tested or mathematically defined than gravity. TWCF has identified a half-dozen leading consciousness theories and intends to eventually fund research to test them all. "What we hope for is a process that reduces the number of incorrect theories," says foundation President Andrew Serazin, who is based in Nassau, Bahamas. "We want to reward people who are courageous in their work, and part of having courage is having the humility to change your mind."

The first two contenders are the global workspace theory (GWT), championed by Stanislas Dehaene of the Collège de France in Paris, and the integrated information theory (IIT), proposed by Giulio Tononi of the University of Wisconsin in Madison. The GWT says the brain's prefrontal cortex, which controls higher order cognitive processes like decision-making, acts as a central computer that collects and prioritizes information from

sensory input. It then broadcasts the information to other parts of the brain that carry out tasks. Dehaene thinks this selection process is what we perceive as consciousness. By contrast, the IIT proposes that consciousness arises from the interconnectedness of brain networks. The more neurons interact with one another, the more a being feels conscious—even without sensory input. IIT proponents suspect this process occurs in the back of the brain, where neurons connect in a gridlike structure.

To test the schemes, six labs will run experiments with a total of more than 500 participants, costing the foundation \$5 million. The labs, in the United States, Germany, the United Kingdom, and China, will use three techniques to record brain activity as volunteers perform consciousness-related tasks: functional magnetic resonance imaging, electroencephalography, and electrocorticography (a form of EEG done during brain surgery, in which electrodes are placed directly on the brain). In one experiment, researchers will measure the brain's response when a person becomes aware of an image. The GWT predicts the front of the brain will suddenly become active, whereas the IIT says the back of the brain will be consistently active.

Tononi and Dehaene have agreed to parameters for the experiments and have registered their predictions. To avoid conflicts of interest, the scientists will neither collect nor interpret the data. If the results appear to disprove one theory, each has agreed to admit he was wrong—at least to some extent.

The project has drawn criticism, mostly because it includes the IIT. Anil Seth, a neuroscientist at the University of Sussex in Brighton, U.K., says the theory is too philosophical—attempting to explain why consciousness exists, rather than how the brain determines whether a stimulus is worthy of conscious attention—to be directly testable. "I don't think [the competition] will do what it says on the tin," he says. Tononi himself doubts the experiments could rule out all aspects of his theory, although he says they could "make our life more difficult."

Koch, who supports the IIT, disagrees with Seth's assessment. "Of course it's testable," he says. He points to clinical trials in which researchers stimulate the brains of people in minimally conscious states such as comas or under anesthesia, triggering brain activity patterns that suggest neurons interconnect in the way the IIT predicts.

Despite his misgivings about the project's prospect for a decisive answer, Seth says it will spark discussion and collaboration among scientific rivals. "That itself is to be applauded," he says. ■

Sara Reardon is a journalist in Bozeman, Montana.

SCIENCE FUNDING

Texas cancer agency seeks new vote of approval

Ballot initiative for CPRIT would raise another \$3 billion to fuel state's cancer research

By Jocelyn Kaiser

Seven years ago, when Francesca Cole was ending her postdoc at Memorial Sloan Kettering Cancer Center in New York City, she got a Texas-size offer: a \$3 million startup package to join the University of Texas's (UT's) MD Anderson Cancer Center in Houston. Most of that sum came from the \$3 billion Cancer Prevention & Research Institute of Texas (CPRIT) in Austin, an unusual state initiative that voters approved in 2007.

On 5 November, Texas residents will decide whether to sustain CPRIT, the second largest public source of cancer funding in the United States after the federal government. At stake is its generous support for 123 tenure-track faculty like Cole, who investigates cancer and DNA repair. Most scientists use fruit flies or zebra fish, but she could afford to build a large team that probes DNA repair using more sophisticated—and expensive—animal models: 25 different genetically modified mice strains. “The CPRIT investment made a huge difference to my research success,” says Cole, who has since won a prestigious New Innovator Award from the National Institutes of Health.

Texas residents appear likely to approve the ballot initiative, which would give CPRIT another \$3 billion through bond sales; a recent poll found that two-thirds of voters support it. Yet some dissent remains from fiscal conservatives. State Senator Charles Schwertner (R) told the *Austin American-Statesman* in January that although CPRIT's goals are “unquestionably noble,” funding cancer research is not a role for state government. He introduced a bill to have CPRIT become a self-sufficient agency, but it failed to advance.

Still, CPRIT and its supporters aren't letting up on their message that the agency has been a boon for Texas's research efforts and the state's economy. “The investment CPRIT has made has accelerated cancer research and the development of the cancer ecosystem in Texas beyond our expectations,” says the agency's chief scientific

officer, James Willson. Geneticist Richard Kolodner of the University of California, San Diego, who chairs CPRIT's external scientific review council, says the initiative will yield an even bigger payoff in the future. “Ten years is a short period of time for a scientific program.”

California's \$3 billion stem cell agency was CPRIT's original inspiration, although California was responding to federal restrictions on stem cell studies, whereas Texas wanted to bolster the state's research efforts and its biotech industry. After a smooth first few years, a scandal broke out in 2012 over a \$18 million incubator award to MD Anderson that had not undergone scientific peer

review. They, in turn, have raised more than \$3 billion from investors. A recent analysis commissioned by CPRIT concluded that the money it pumps into the economy generates \$1.4 billion in annual economic activity and supports 10,000 jobs. And one “immediate” result of CPRIT's \$250 million in prevention grants has been cancer screening and other services for 320,000 Texans a year, Willson says.

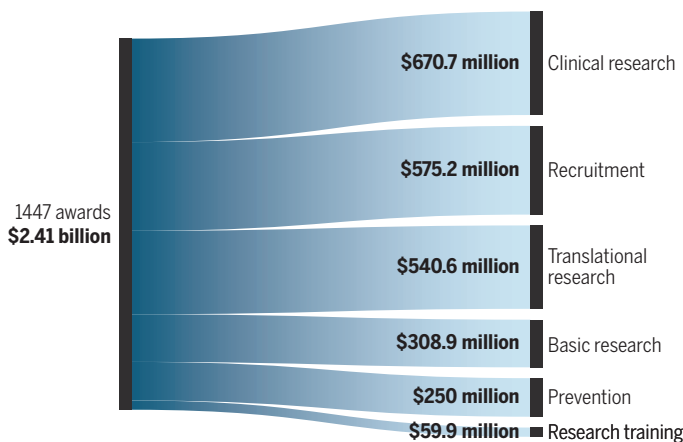
Kolodner highlights CPRIT's support for recruiting 181 researchers as “the gift that keeps on giving.” They include heavyweights such as MD Anderson's Jim Allison, who won a Nobel Prize last year for his work on immunotherapies, and stem cell researcher Sean Morrison of UT's Southwestern Medical Center in Dallas. But one large investment misfired: Mouse geneticists Neal Copeland and Nancy Jenkins, recruited in 2011 from Singapore to the Methodist Hospital Research Institute in Houston with a 5-year, \$29 million joint package—half from CPRIT—retired after that money ran out. The agency has since capped its share of such awards to \$6 million in order to recruit more scientists.

CPRIT says its recruits have drawn additional grant money from other sources. Yet Texas's National Cancer Institute funding, which totaled \$249 million in 2018, has risen only 16% since 2008, below NCI's overall budget growth of 23% over the same decade. Willson counters that since CPRIT's inception, NCI has tabbed two Texas sites as new Comprehensive Cancer Centers, which “catalyze something much greater than sum of the parts.” (MD Anderson was already a center.) Willson adds that many early-career CPRIT scholars are just now getting their first NCI grants, and he says funding from private research foundations has ballooned.

CPRIT may still be an experiment, but Texas voters will almost certainly give it a chance to continue. That stands in contrast to California, where efforts to renew the California Institute for Regenerative Medicine face an uphill battle. Kolodner says, “I think sadly people probably relate to cancer much more than stem cells.” ■

State of cancer funding

Texas's cancer funding agency has handed out most of its initial \$3 billion, boosting prevention services, academic research, and product development.



review. That, along with concerns that politics was skewing grant decisions, prompted CPRIT's chief scientific officer, Nobel laureate and biochemist Alfred Gilman, to resign in protest, along with most of its scientific council and many grant reviewers. After more problems led to a 10-month hold on new grants and a governance overhaul, the agency got back on track.

CPRIT has awarded more than \$2.4 billion for 1447 awards split among clinical and translational research, recruitment, basic research, training, and prevention. It has supported shared resources such as bioinformatics facilities and advanced microscopes. The agency touts its practical impact, saying 36 cancer companies have used its money to launch, grow, or move to the

OUTSIZE

Pint-size proteins, found everywhere from venom to muscles, can block or boost their bigger kin *By Mitch Leslie*

Mice put human runners to shame. Despite taking puny strides, the rodents can log 10 kilometers or more per night on an exercise wheel. But the mice that muscle biologist Eric Olson of the University of Texas Southwestern Medical Center in Dallas and colleagues unveiled in 2015 stood out. On a treadmill, the mice could scurry up a steep 10% grade for about 90 minutes before faltering, 31% longer than other rodents. Those iron mice differed from counterparts in just one small way—the researchers had genetically altered the animals to lack one muscle protein. That was enough to unleash superior muscle performance. “It’s like you’ve taken the brakes off,” Olson says.

Just as startling was the nature of the crucial protein. Muscles house some gargantuan proteins. Dystrophin, a structural protein whose gene can carry mutations that cause muscular dystrophy, has more than 3600 amino acids. Titin, which acts like a spring to give muscles elasticity, is the biggest known protein, with more than 34,000 amino acids. The protein disabled in the mice has a paltry 46. Although researchers have probed how muscles work for more than 150 years, they had completely missed the huge impact this tiny protein, called myoregulin, has on muscle function.

Olson and his colleagues weren’t the only ones to be blindsided by Lilliputian proteins. As scientists now realize, their initial rules for analyzing genomes discriminated against identifying those pint-size molecules. Now, broader criteria and better detection

methods are uncovering minuscule proteins by the thousands, not just in mice, but in many other species, including humans. “For the first time, we are about to explore this universe of new proteins,” says biochemist Jonathan Weissman of the University of California, San Francisco.

Biologists are just beginning to delve into the functions of those molecules, called microproteins, micropeptides, or miniproteins. But their small size seems to allow them to jam the intricate workings of larger proteins, inhibiting some cellular processes while unleashing others. Early findings suggest microproteins bolster the immune system, control destruction of faulty RNA molecules, protect bacteria from heat and cold, dictate when plants flower, and provide the toxic punch for many types of venom. “There’s probably going to be small [proteins] involved in all

biological processes. We just haven’t looked for them before,” says biochemist Alan Saghatelian of the Salk Institute for Biological Studies in San Diego, California.

Small proteins also promise to revise the current understanding of the genome. Many appear to be encoded in stretches of DNA—and RNA—that were not thought to help build proteins of any sort. Some researchers speculate that the short stretches of DNA could be newborn genes, on their way to evolving into larger genes that make full-size proteins. Thanks in part to small proteins, “We need to rethink what genes are,” says microbiologist and molecular biologist Gisela Storz of the National Institute of Child Health and Human Development in Bethesda, Maryland.

Despite the remaining mysteries, scientists are already testing potential uses for the molecules. One company sells insecticides derived from small proteins in the poison of an Australian funnel-web spider. And a clinical trial is evaluating an imaging agent based on another minute protein in scorpion venom, designed to highlight the borders of tumors so that surgeons can remove them more precisely. Many drug companies are now searching for small proteins with medical potential, says biochemist Glenn King of the University of Queensland in St. Lucia, Australia. “It’s one of the most rapidly growing areas.”

OTHER SHORT AMINO ACID chains, often called peptides or polypeptides, abound in cells, but they are pared-down remnants of bigger predecessors. Myoregulin and its diminutive brethren, in contrast, are born small. How tiny they can be remains unclear. Fruit flies



The venom of this predatory water bug has more than a dozen small proteins.

SALIVARY AMYLASE
496 AMINO ACIDSMYOREGULIN
46 AMINO ACIDS

IMPACT



Myoregulin is small compared with a typical protein and minute next to a behemoth such as titin.

rely on a microprotein with 11 amino acids to grow normal legs, and some microbes may crank out proteins less than 10 amino acids long, notes microbial genomicist Ami Bhatt of Stanford University in Palo Alto, California. But even the largest small proteins don't measure up to average-size proteins such as alpha amylase, a 496-amino-acid enzyme in our saliva that breaks down starch.

Few small proteins came to light until recently because of a criterion for identifying genes set about 20 years ago. When scientists analyze an organism's genome, they often scan for open reading frames (ORFs), which are DNA sequences demarcated by signals that tell the cell's ribosomes, its protein-making assembly lines, where to start and stop. In part to avoid a data deluge, past researchers typically excluded any ORF that would yield a protein smaller than 100 amino acids in eukaryotes or 50 amino acids in bacteria. In yeast, for example, that cutoff limited the list of ORFs to about 6000.

Relaxing that criterion reveals that cells carry vastly more ORFs. Earlier this year, Stanford postdoc Hila Sberro Livnat, Bhatt, and colleagues trawled genome fragments from the microbes that inhabit four parts of the human body, including the gut and skin. By searching for small ORFs that could encode proteins between five and 50 amino acids long, the researchers identified about 4000 families of potential microproteins. Almost half resemble no known proteins, but the sequence for one small ORF suggested that a corresponding protein resides in ribosomes—a hint that it could play some fundamental role. "It's not just genes with es-

oteric functions that have been missed" when scientists overlooked small ORFs, Bhatt says. "It's genes with core functions."

Other cells also house huge numbers of short ORFs—yeast could make more than 260,000 molecules with between two and 99 amino acids, for example. But cells almost certainly don't use all those ORFs, and some of the amino acid strings they produce may not be functional. In 2011, after finding more than 600,000 short ORFs in the fruit fly genome, developmental geneticist Juan Pablo Couso of the University of Sussex in Brighton, U.K., and colleagues tried to whittle down the number. They reasoned that if a particular ORF had an identical or near-identical copy in a related species, it was less likely to be genomic trash. After searching another fruit fly's genome and analyzing other evidence that the sequences were being translated, the group ended up with a more manageable figure of 401 short ORFs likely to yield microproteins. That would still represent a significant fraction of the insects' protein repertoire—they harbor about 22,000 full-size proteins.

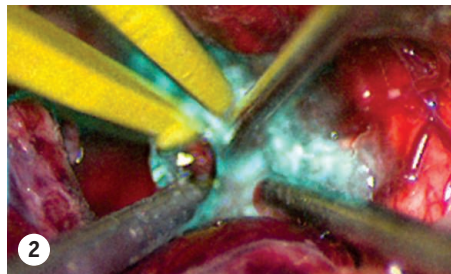
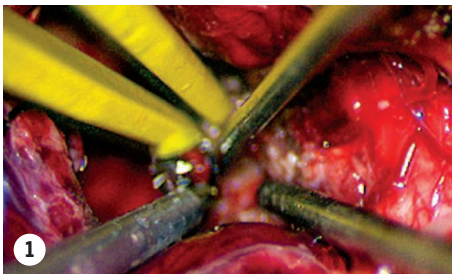
Weissman and colleagues found microproteins a second way, through a method they invented to broadly determine which proteins cells are making. To fashion any protein, a cell first copies a gene into messenger RNA. Then ribosomes read the mRNA and string together amino acids in the order it specifies. By sequencing mRNAs attached to ribosomes, Weissman and his team pinpoint which ones cells are actually turning into proteins and where on the RNAs a ribosome starts to read. In a 2011 *Cell* study, he and his team applied that ribosome profiling method, also called Ribo-seq, to mouse embryonic stem cells and discovered the cells were making thousands of unexpected proteins, including many that would fall below the 100-amino-acid cutoff. "It was quite clear that the standard understanding had ignored a large universe of proteins, many of which were short," Weissman says.

Saghatelyan and his colleagues adopted a third approach to discover a trove of microproteins in our own cells. The researchers used mass spectrometry, which involves breaking up proteins into pieces that are



A microprotein in the poison of the deathstalker scorpion has been fused to a fluorescent dye to make tumors emit near-infrared light.

- 1 A tumor seen in visible light
- 2 Same tumor in visible and near-infrared light



sorted by mass to produce a distinctive spectrum for each protein. Saghatelian, his then-postdoc Sarah Slavoff, and colleagues applied the method to protein mixtures from human cells and then subtracted the signatures of known proteins. That approach revealed spectra for 86 previously undiscovered tiny proteins, the smallest just 18 amino acids long, the researchers reported in 2013 in *Nature Chemical Biology*.

BEING SMALL LIMITS a protein's capabilities. Larger proteins fold into complex shapes suited for a particular function, such as cata-

lyzing chemical reactions. Proteins smaller than about 50 to 60 amino acids probably don't fold, says chemist Julio Camarero of the University of Southern California in Los Angeles. So they probably aren't suited to be enzymes or structural proteins.

However, their diminutive size also opens up opportunities. "They are tiny enough to fit into nooks and crannies of larger proteins that function as channels and receptors," Olson says. Small proteins often share short stretches of amino acids with their larger partners and can therefore bind to and alter the activity of those proteins. Bound micro-

proteins can also shepherd bigger molecules to new locations—helping them slip into cell membranes, for instance.

Because of their attraction to larger proteins, small proteins may give cells a reversible way to switch larger proteins on or off. In a 2016 study in *PLOS Genetics*, plant developmental biologist Stephan Wenkel of the University of Copenhagen and colleagues genetically altered *Arabidopsis* plants to produce extra amounts of two small proteins. The plants normally burst into flower when the days are long enough, but when they overproduced the two microproteins, their flowering was postponed. The small proteins caused that delay by blocking a hefty protein called CONSTANS that triggers flowering. They tether CONSTANS to other inhibitory proteins that shut it down. "A cell uses things that help it survive. If a short protein does the job, that's fine," Saghatelian says.

Those jobs include other key tasks. In 2016, Slavoff, Saghatelian, and colleagues revealed that human cells manufacture a 68-amino-acid protein they named NoBody that may help manage destruction of faulty or unneeded mRNA molecules. NoBody's name reflects its role in preventing formation of processing bodies (P-bodies), mysterious clusters in the cytoplasm where RNA breakdown may occur. When the protein is missing, more P-bodies form, thus boosting RNA destruction and altering the cell's internal structure. "It shows that small proteins can have massive effects in the cell," Slavoff says.

Muscles appear to depend on a variety of microproteins. During embryonic development, individual muscle cells merge into fibers that power contraction. The 84-amino-acid protein myomixer teams up with a larger protein to bring the cells together, Olson's team reported in 2017 in *Science*. Without it, embryonic mice can't form muscles and are almost transparent.

Later in life, myoregulin steps in to help regulate muscle activity. When a muscle receives a stimulus, cellular storage depots spill calcium, triggering the fibers to contract and generate force. An ion pump called SERCA then starts to return the calcium to storage, allowing the muscle fibers to relax. Myoregulin binds to and inhibits SERCA, Olson's team found. The effect limits how often a mouse's muscles can contract—perhaps ensuring

that the animal has muscle power in reserve for an emergency, such as escaping a predator. Another small protein, DWORF, has the opposite effect, unleashing SERCA and enabling the muscle to contract repeatedly.

Even extensively studied organisms such as the intestinal bacterium *Escherichia coli* harbor unexpected small proteins that have important functions. Storz and her team reported in 2012 that a previously undiscovered 49-amino-acid protein called AcrZ helps the microbe survive some antibiotics by stimulating a pump that expels the drugs.

And the venom produced by a variety of organisms—including spiders, centipedes, scorpions, and poisonous mollusks—teems with tiny proteins. Many venom components disable or kill by blocking the channels for sodium or other ions that are necessary for transmission of nerve impulses. Small proteins “hit these ion channels with amazing specificity and potency,” King says. “They are the major components of venoms and are responsible for most of the pharmacological and biological effects.”

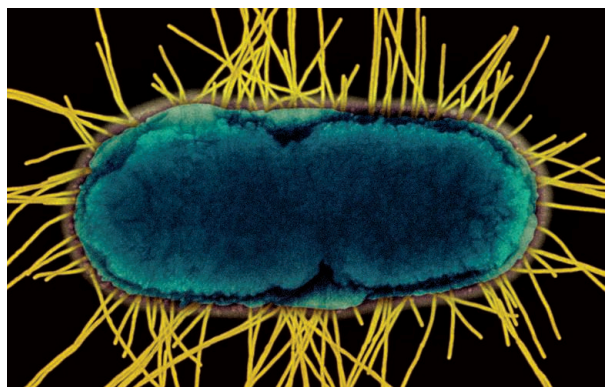
Australia’s giant fish-killing water bug (see p. 296), for instance, doesn’t just rely on sharp claws and lancelike mouthparts to subdue prey. It injects its victims with a brew of more than 130 proteins, 15 of which have fewer than 100 amino acids, King and colleagues reported last year.

UNLIKE HULKING PROTEINS such as antibodies, microproteins delivered by pill or injection may be able to slip into cells and alter their functions. Captopril, the first of a class of drugs for high blood pressure known as angiotensin-converting enzyme inhibitors was developed from a small protein in the venom of a Brazilian pit viper. But the drug, which the Food and Drug Administration approved for sale in the United States in 1981, was discovered by chance, before scientists recognized small proteins as a distinct group. So far, only a few microproteins have reached the market or clinical trials.

Cancer researchers are trying to capitalize on a microprotein in the poison of the death-stalker scorpion (*Leiurus quinquestriatus*) of Africa and the Middle East. The molecule has

a mysterious attraction to tumors. By fusing it to a fluorescent dye, scientists hope to illuminate the borders of brain tumors so that surgeons can safely cut out the cancerous tissue. “It lights up the tumor. You can see the margins and if there are any metastases,” King says. A clinical trial is now evaluating whether the dual molecule can help surgeons remove brain tumors in children.

How important small proteins will be for medicine is still unknown, but they have already upended several biological assumptions. Geneticist Norbert Hübner of the Max Delbrück Center for Molecular Medi-



Bacteria such as *Escherichia coli* also churn out many microproteins, although their functions remain unclear in many cases.

cine in Berlin and colleagues found dozens of new microproteins in human heart cells. The group traced them to an unexpected source: short sequences within long noncoding RNAs, a variety that was thought not to produce proteins. After identifying 169 long noncoding RNAs that were probably being read by ribosomes, Hübner and his team used a type of mass spectrometry to confirm that more than half of them yielded microproteins in heart cells, a result reported earlier this year in *Cell*.

The DNA sequences for other tiny proteins also occur in unconventional locations. For example, some lie near the ORFs for bigger proteins. Researchers previously thought those sequences helped manage the production of the larger proteins, but rarely gave rise to proteins themselves. Some coding sequences for recently discovered microproteins are even nested within sequences that encode other, longer proteins.

Those genomic surprises could illuminate how new genes arise, says evolutionary systems biologist Anne-Ruxandra Carvunis of the University of Pittsburgh in Pennsylvania. Researchers had thought most new genes emerge when existing genes duplicate or fuse, or when species swap DNA. But to Carvunis, microproteins suggest protogenes can form when mutations create new start and stop signals in a noncoding portion of the genome. If the resulting ORF produces a beneficial protein, the novel sequences would remain in the genome and undergo natural selection, eventually evolving into larger genes that code for more complex proteins.

In a 2012 study, Carvunis, who was then a postdoc in the lab of Marc Vidal at the Dana-Farber Cancer Institute in Boston, and colleagues found that yeast translate more than 1000 short ORFs into proteins, implying that these sequences are protogenes. In a new study, Carvunis and her team tested whether young ORFs can be advantageous for cells. They genetically altered yeast to boost output of 285 recently evolved ORFs, most of which code for molecules that are smaller the standard protein cutoff or just over it. For almost 10% of the proteins, increasing their levels enhanced cell growth in at least one environment. The results, posted on the preprint server bioRxiv, suggest these sequences could be on their way to becoming full-fledged genes, Carvunis says.

Slavoff still recalls being astonished when, during her interview for a postdoc position with Saghatelian, he asked whether she would be willing to go hunting for small proteins. “I had never thought that there could be this whole size of proteins that was dark to us until then.”

But the bet paid off—she now runs her own lab that is searching for microproteins. Recently, she unleashed some of her postdocs and graduate students on one of the most studied organisms, the K12 strain of *E. coli*. The team soon uncovered five new microproteins. “We are probably only scratching the surface,” she says. ■

INSIGHTS

PERSPECTIVES

EVOLUTION

Evolutionary history of tissue bending

Newly discovered unicellular eukaryote shows hallmark feature of animal morphogenesis

By Pavel Tomancak

Animal embryos shape their tissues during development through a variety of mechanisms, one of which involves coordinated constriction of one side of a sheet of cells, leading to tissue bending. This apical constriction mechanism of tissue morphogenesis occurs in most animal groups (1), suggesting that it was inherited from a common ancestor. On page 326 of this issue, Brunet *et al.* (2) describe the remarkable discovery of a new species of colony-forming unicellular eukaryote that uses collective cell contraction to change their morphology and behavior in response to a lack of light. Because the new species belongs to the closest living relatives of animals, the choanoflagellates (3), these results cast new light on the evolutionary origin of collective cell contractility.

In a pond above the tidal line on a beach in the Caribbean island of Curaçao, Brunet *et al.* found a remarkable creature composed of ~100 cells with whip-like appendages (flagella) that together formed a small sheet. The sheet was bent inward, with the flagella pointing toward the interior of a cup-shaped colony. However, occasionally the sheet would invert its curvature rapidly, and within 30 s all the flagella would be facing outward along the radius of the inverted colony (see the image). This change in tissue morphology is similar to developmental morphogenesis in complex multicellular animals, in which epithelial sheets fold and invaginate to form multilayered structures.

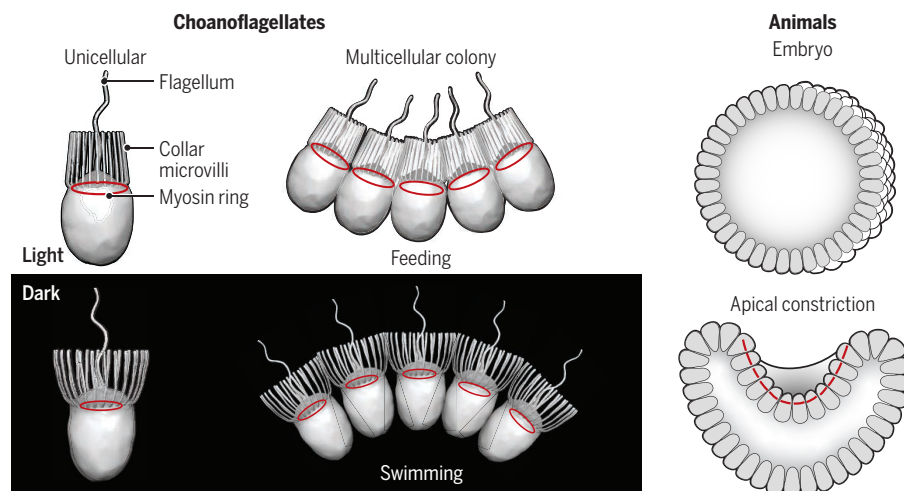
Brunet *et al.* studied the creature and on the basis of the morphology of the cells making up the sheet concluded that it was a choanoflagellate. Many researchers study these unicellular eukaryotes that sometimes form multicellular colonies (4), owing to their distinct evolutionary relatedness to animals. What Brunet *et al.* found was a new species of choanoflagellate, named by the authors *Choanoeca flexa*. The inversion behavior is specific to *C. flexa* but has been observed before. In 1983, a study of the closely related *Choanoeca perplexa* mentions such inversion behavior (5); however, that colony

did not thrive in the laboratory and was lost and never studied again. Brunet *et al.* managed to keep their *C. flexa* alive in the laboratory by growing it together with the bacteria it fed on in the pond in which it was discovered. What followed is a triumph of unleashing the power of modern biology on a newly discovered species.

The authors found that the inversion behavior is triggered by switching off the light. Analyses of the *C. flexa* genome revealed a gene, encoding the protein rhodopsin phosphodiesterase (6), that could both sense light and pass the information

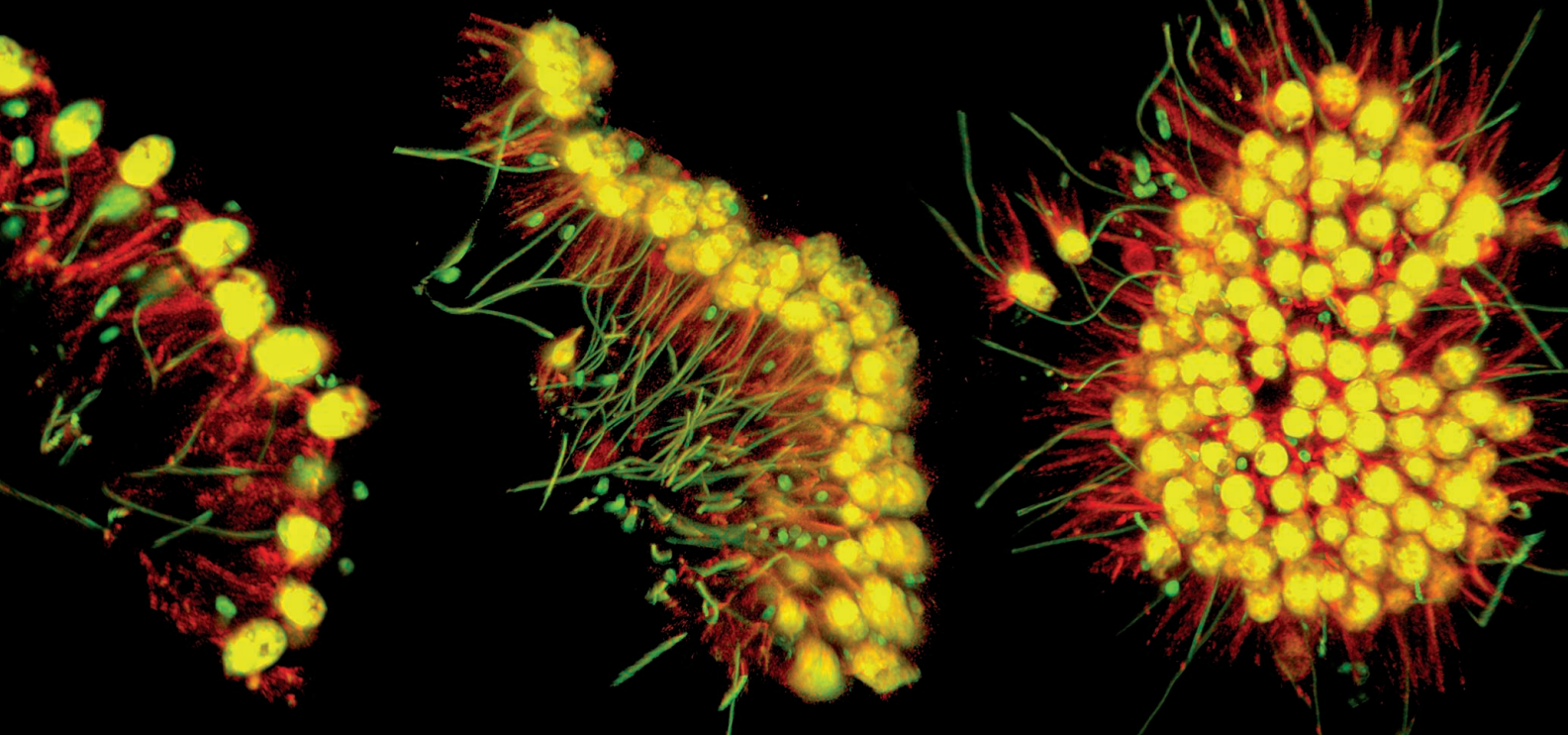
Myosin-mediated morphogenesis in choanoflagellates and animals

In choanoflagellates, single cells or sheets of connected cells contract because of actomyosin activity. This behavior is triggered by darkness and results in a change in morphology (opening of collar microvilli and sheet inversion) and behavior (from feeding to swimming). In animal embryos, local myosin-induced apical constriction leads to epithelial tissue invagination during gastrulation.



Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. Email: tomancak@mpi-cbg.de

GRAPHIC: N. DESAI/SCIENCE; PHOTO: THIBAUT BRUNET



to other cellular processes by modifying cyclic nucleotides that are commonly used in intracellular signaling. This gene was a good candidate to mediate the observed light-dependent behavior, but how could this be proven in a newly discovered species in which gene expression cannot be manipulated? Brunet *et al.* designed an elegant experiment to address this question.

Photodetection with rhodopsin requires a cofactor, a chromophore called retinal (7). *C. flexa* does not have the genes necessary to produce retinal. Therefore, it must be getting retinal from food, the bacteria. The authors fed their choanoflagellates with bacteria that can or cannot produce retinal and found that only colonies fed on retinal-producing bacterial strains can invert in darkness. Moreover, addition of synthetic retinal alone triggered the inversion. But what does inversion do? The authors showed that when the cell colony has its flagella facing inward, it can eat better by capturing the bacteria from the surrounding medium. However, when it inverts, the flagella are oriented in a way that is more useful for moving the colony around, potentially to escape a large predator casting a shadow over the colony. Therefore, the sheet inversion mediates a trade-off between feeding and swimming.

Having established the regulation and ecological relevance of the inversion behavior, Brunet *et al.* examined the mechanism of inversion. They noticed from microscopy observations that the individual cells of *C. flexa* hold together by means of contacts between cellular protrusions surrounding the flagellum that are called collar microvilli. This is a very specific way of holding a sheet of cells together, quite distinct from how epithe-

lial sheets of animal cells are connected by junctional complexes. However, the *C. flexa* shares a remarkable molecular feature with animal epithelia: At the base of the collar, there is a prominent actomyosin ring, a cytoskeletal feature that is associated with cellular contractility in animal cells (8). During inversion, the ring contracts and the collar opens up. Chemical inhibitors of actomyosin contractility abolish this behavior. Therefore, the inversion behavior of choanoflagellates uses the same actomyosin molecular machinery deployed by animal cells to sculpt tissues during development through apical constriction (see the figure).

Myosin protein sequences are highly similar between choanoflagellates and animals (9), and therefore, the myriad of molecular reagents developed in laboratories studying myosin in animal cells can be used in choanoflagellates. Brunet *et al.* used these reagents to visualize myosin distribution in several species across the choanoflagellate evolutionary tree and found that they all have the myosin rings observed in *C. flexa*.

Therefore, the study of Brunet *et al.* indicates that contractile actomyosin machinery was inherited by choanoflagellates and animals from a common ancestor. Animals have evolved ways to use this contractile machinery to bend epithelial tissues by means of apical constriction—for example, during gastrulation. In choanoflagellates such as *C. flexa*, the machinery mediates collective cell inversion behavior in cell colonies. In animals, the contractility is regulated through mechanochemical signals and feedback during morphogenesis (10), whereas in choanoflagellates, the contractility is directly linked to environmental stimuli such as absence

The successive steps of the inversion process.

The flagella are at the apical pole of each cell (green) where the collar of microvilli (red), which drive inversion, are also located.

of light and the decisions between feeding and predator-escaping motile behaviors. Actomyosin contractility is therefore an ancient morphogenetic module (11) that has been, most likely independently, co-opted to achieve diverse morphogenetic outcomes under vastly different regulation in animals and choanoflagellates.

The new species was discovered only a year ago, and yet Brunet *et al.* were able to convincingly connect the molecular mechanisms of actomyosin contraction to the collective inversion behavior and its ecological importance. This shows that comprehensive mechanistic studies are no longer confined to established organismal models such as fruit-fly, zebrafish, or mouse. On the contrary, the study by Brunet *et al.* highlights the power of looking at biodiversity to uncover new biological principles. Their approach of mapping fundamental morphogenetic modules across the tree of life promises future insight into how organismal shapes evolve. ■

REFERENCES AND NOTES

1. A. C. Martin, B. Goldstein, *Development* **141**, 1987 (2014).
2. T. Brunet *et al.*, *Science* **366**, 326 (2019).
3. N. King *et al.*, *Nature* **451**, 783 (2008).
4. B. S. C. Leadbeater, *The Choanoflagellates: Evolution, Biology and Ecology* (Cambridge Univ. Press, 2015).
5. B. S. C. Leadbeater, *J. Mar. Biol. Assoc.* **63**, 135 (1983).
6. K. Yoshida *et al.*, *J. Biol. Chem.* **292**, 7531 (2017).
7. O. P. Ernst *et al.*, *Chem. Rev.* **114**, 126 (2014).
8. J. M. Sawyer *et al.*, *Dev. Biol.* **341**, 5 (2010).
9. T. A. Richards, T. Cavalier-Smith, *Nature* **436**, 1113 (2005).
10. E. Hannezo, C.-P. Heisenberg, *Cell* **178**, 12 (2019).
11. S. A. Newman, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **371**, 20150443 (2016).

10.1126/science.aaz1289

IMMUNOTHERAPY

Mobilizing unconventional T cells

Modulating unconventional antigen presentation could treat infections and cancer

By **Tom H. M. Ottenhoff** and **Simone A. Joosten**

The human leukocyte antigen-E (HLA-E) molecule is expressed on all nucleated cells and presents self-peptides. The resulting complexes can bind to inhibitory receptors on immune killer cells, including ~50% of natural killer (NK) cells and a subset (~5%) of CD8⁺ cytotoxic T lymphocytes (CTLs). Interruption of this inhibitory axis, which serves as an immune checkpoint, can improve effector functions of both CTLs and NK cells and enhance antitumor activity (1). HLA-E antigen expression can also be used as a mechanism for pathogen-infected cells to avoid being killed by NK cells. However, HLA-E can present peptides from pathogens and tumor cells to so-called unconventional CD8⁺ T cells, which can then be mobilized to fight the infection or tumor. Recent advances in understanding this dichotomy between unconventional T cell activation and NK cell suppression reveal potential preventive and therapeutic applications in infectious diseases and cancer.

HLA-E is an HLA class Ib molecule and is classified as nonclassical because it is relatively monomorphic. In contrast to classical HLA class Ia molecules, which are highly polymorphic, the *HLA-E* locus encodes only two functional variants, HLA-E*0101 and HLA-E*0103. These two proteins differ by one amino acid (Arg¹⁰⁷ or Gly¹⁰⁷, respectively), which is located outside the HLA-E peptide binding groove and is therefore considered unlikely to influence HLA-E-peptide interactions directly (2, 3). Although largely similar, there are slight differences in bound-peptide repertoires, peptide-binding affinities, the level of expression, and the stability of both HLA-E molecules for reasons that are poorly understood (2). Both variants are maintained at comparable frequencies in the human population, potentially reflecting a lack of selective advantage for either variant.

HLA class Ia expression is often down-regulated in human tumors, which facilitates immune evasion from classical CD8⁺ CTL-mediated killing. However, HLA class Ia down-regulation can make tumors sus-

ceptible to NK cell-mediated killing because HLA class Ia is the main supplier of peptides that can bind to, and thus maintain, cell surface expression of HLA-E. It has become clear, though, that when HLA class Ia is absent, new tumor peptides can be presented by HLA-E to activate unconventional CD8⁺ T cells through HLA-E-peptide-specific recognition by T cell receptors (TCRs). There are several types of unconventional T cells, which recognize antigens through TCRs by means of monomorphic antigen-presentation molecules (3). In the case of HLA-E, the unconventional CD8⁺ T cells reported thus far can have both cytotoxic and suppressive properties (4, 5). In mice, Qa-1 (the mouse ortholog of HLA-E)-restricted unconventional CD8⁺ T cells can confer antitumor activity and improve survival (6, 7). These findings inspired the search for potential peptides in other diseases, including those caused by persistent pathogen infections, such as human immunodeficiency virus (HIV), cytomegalovirus (CMV), and *Mycobacterium tuberculosis* (Mtb). This resulted in the identification of multiple peptides that are recognized by unconventional HLA-E-restricted CD8⁺ T cells and that were derived from Mtb, Epstein Barr virus (EBV), HIV, and *Salmonella typhi* (2, 3), highlighting the importance of HLA-E antigen presentation in the activation of unconventional CD8⁺ T cells.

Peptides from Mtb, HIV, and CMV in complex with HLA-E*0103 were recently crystallized. The resulting structures showed canonical anchoring of peptides into the peptide-binding groove (8) and revealed how peptides might be designed to activate HLA-E-restricted T cells for vaccination. Three TCR molecules have been co-crystallized with their cognate HLA-E-bound peptide, revealing canonical contact sites (9). Further investigation is required to understand which structural features restrict TCRs to HLA-E-peptide complexes.

Recent studies in nonhuman primates (NHPs) revealed the importance of HLA-E-restricted T cells in immune responses to infectious diseases. Simian immunodeficiency virus (SIV; the NHP equivalent to HIV) antigens were expressed in modified rhesus CMV (RhCMV) vectors as a subunit-vaccination strategy, because CMV viruses

are known to induce strong and long-lasting CD8⁺ memory T cell responses (2). A subunit vaccine elicits immunological memory through exposure to selected immunogenic components from a tumor or pathogen. Administration of this RhCMV-SIV subunit vaccine to NHPs resulted in protection against subsequent SIV infection, which was mediated through unconventional CD8⁺ T cells that recognized SIV peptides bound to either MHC class II molecules (which are expressed by antigen-presenting cells and present peptides from extracellular proteins) or MHC-E (the ortholog of HLA-E), but not conventional MHC class Ia molecules. In over half of the animals, this resulted in eradication of experimentally induced SIV infection (10). Similarly designed RhCMV-TB-antigen vectors also induced complete protection against experimental tuberculosis in 41% of treated NHPs. In these animals, equivalent protection could also be achieved with vectors that induced conventional CD8⁺ T cells and CD4⁺ T helper cells, suggesting redundancy in unconventional and conventional CD8⁺ T cell responses in the NHP-tuberculosis model (11). Regardless of the many unresolved questions in HLA-E biology, these data collectively support the candidacy of HLA-E as a targetable pathway for vaccination as well as immunotherapy—for example, by antibody-mediated blockade of CD8⁺ CTL or NK cell-expressed inhibitory receptor molecules, one of which is NK group 2A (NKG2A) (1). In addition to the relative monomorphism of HLA-E, an advantage of HLA-E-based vaccines over traditional vaccine strategies targeting HLA class Ia molecules is that HLA-E expression is not down-regulated when HIV and Mtb infection co-occur, which is an important global health issue.

There is much more to be understood about the immunology of HLA-E and how this can be translated into vaccines and immunotherapies. There is limited understanding of which peptides are optimal targets for HLA-E-restricted T cells. Developing new tools to identify these from pathogen—and tumor—genomic sequences would be valuable. Moreover, exactly how and where peptide antigens are processed intracellularly for HLA-E presentation is largely unknown. Studies in mice suggest a typical endoplasmic reticulum peptide-loading pathway for Qa-1 (7). However, in the case of Mtb infection, and likely other intracellular pathogens, HLA-E can be expressed in the phagosome, suggesting an alternative site of HLA-E peptide loading (3, 7). Understanding the biology of HLA-E antigen presentation will be key to the design of optimal strategies to target this

Department of Infectious Diseases, Leiden University Medical Center, Leiden, the Netherlands.
Email: t.h.m.ottenhoff@lumc.nl

pathway for unconventional CD8⁺ T cell activation.

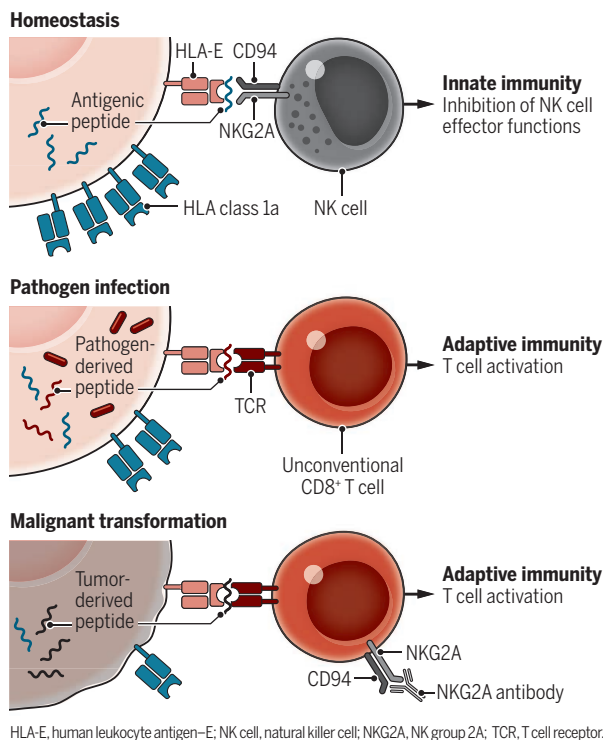
Another avenue for further investigation is the diversity of the TCR repertoire for HLA-E-presented ligands. For HLA-E-CMV peptides, preferential usage of the TCR β -chain variable region (V β 16) has been reported, and mouse tumor models suggest a role for semi-invariant TCR $\alpha\beta$ in recognizing certain self-peptides bound to Qa-1 (12, 13). An unanswered question is whether HLA-E-restricted unconventional T cells display narrow TCR repertoires similar to those in the mouse, or broader TCR repertoires, and whether these differ according to disease. It also remains unclear how the HLA-E-restricted TCR repertoire is selected in naïve T cells in the thymus.

For successful vaccine or immunotherapy development, it will be critical to demonstrate that identified HLA-E peptide ligands are expressed at the surface of tumor or infected cells at densities and durations that are sufficient to engage TCRs and induce T cell activation and thus immunological memory that is important in vaccine responses. This should include analysis of ligand expression in the affected organ—for example, the lungs for tuberculosis (14). Furthermore, little is known about the durability and memory capacity of HLA-E-restricted T cells. In mouse tumor models, peptide vaccination could induce memory CD8⁺ T cells specific for Qa-1-restricted tumor peptides (13). Data from the RhCMV studies in NHPs suggest potent, long-term induction of effector memory T cells, but this could also reflect the continuous presence of antigen expressed from replicating RhCMV. In human CMV infection, expansion of memory CD8⁺ T cells is observed (2), probably reflecting the same phenomenon.

Suitable and affordable small-animal models would be of value in exploring HLA-E biology, such as the availability of HLA-E*0101 and HLA-E*0103 transgenic mice. This will allow comparative studies of vaccine formulations using different delivery systems (adjuvanted peptides, viral vectors, bacterial carriers) to optimize vaccine efficacy. In mouse tumor models and the RhCMV-SIV NHP model, the association of MHC-E and Qa-1-restricted T cells with protective immunity suggests an important role for MHC-E and Qa-1, but specific depletion studies have thus far not been performed. Mtb infection of mice genetically lacking Qa-1 resulted in more severe tuberculosis than in wild-type

Human leukocyte antigen-E in immunity

Under homeostatic conditions, HLA-E presents self-peptides and prevents NK cell-mediated lysis through the CD94-NKG2A axis, thus regulating innate immunity. In addition, HLA-E can present pathogen- or tumor-derived peptide antigens to unconventional CD8⁺ T cells, which recognize peptide-HLA-E complexes through specific TCRs, regulating adaptive immunity.



animals, suggesting a protective role for Qa-1-restricted CD8⁺ T cell responses during Mtb infection (15). Consistently, Qa-1-restricted CD8⁺ T cells were cytolytic and could suppress other T cells, a phenotype replicated by HLA-E-restricted human CD8⁺ T cells cultured in vitro (5). Perhaps HLA-E-restricted CD8⁺ T cells contribute to protective immunity to Mtb infection, and likely other pathogens, by simultaneously killing infected cells and inhibiting intracellular infection (as shown for Mtb), while also suppressing inflammation and thereby limiting collateral tissue damage.

The biology of HLA-E is intriguing. For example, why are both alleles maintained, are they redundant, and what controls the unexpected differences between them? Puzzling in this context is the much higher number of functional MHCE variants in NHPs: ~30 variants have been described (2). Additionally, human Mtb-specific HLA-E-restricted T cells were described to possess a T helper 2 (T_H2)-like phenotype, including production of the cytokines interleukin-4 (IL-4), IL-5, and IL-13, and induced B cell activation through IL-4 (5). However, the functional role of this T_H2-like phenotype and the contribution of B cells in control-

ling Mtb infection remain unclear. At which sites do HLA-E-restricted T cells act most prominently—mucosally or systemically? It will also be important to ascertain whether these are tissue-resident memory cells that can be targeted by mucosal vaccination.

Translational research can already begin to harness the knowledge of HLA-E biology to develop new vaccine and immunotherapeutic approaches. Such strategies include preventive or therapeutic subunit vaccines that can mobilize unconventional T cells, or T cells expressing engineered TCRs that recognize peptide-HLA-E complexes on infected or malignant cells. Alternatively, high-affinity soluble TCR molecules can be engineered that target malignant or infected cells with high precision. Vaccines for infectious and malignant diseases may be designed on the basis of relatively small numbers of pathogen- or tumor-derived HLA-E-presented peptides, formulated in suitable adjuvants. In the case of established tumors, T cell-activating vaccines may need strong potentiation—for example, by combination with immune checkpoint-blocking antibodies that prevent NKG2A binding to HLA-E (see the

figure)—to relieve the immune checkpoint on NK and CD8⁺ T cell populations. These HLA-E-centered strategies could help to improve immune control of infectious diseases and cancer (1, 3). ■

REFERENCES AND NOTES

1. P. André et al., *Cell* **175**, 1731 (2018).
2. H. R. Sharpe, G. Bowyer, S. Brackenridge, T. Lambe, *Clin. Exp. Immunol.* **196**, 167 (2019).
3. S. A. Joosten et al., Collaboration for Tuberculosis Vaccine Discovery - Donor-Unrestricted T-cells Working Group, Bill and Melinda Gates Foundation, *Vaccine* **37**, 3022 (2019).
4. S. A. Joosten et al., *PLOS Pathog.* **6**, e1000782 (2010).
5. K. E. van Meijgaarden et al., *PLOS Pathog.* **11**, e1004671 (2015).
6. C. C. Oliveira et al., *J. Exp. Med.* **207**, 207 (2010).
7. N. A. Nagarajan, F. Gonzalez, N. Shastri, *Nat. Immunol.* **13**, 579 (2012).
8. L. C. Walters et al., *Nat. Commun.* **9**, 3137 (2018).
9. L. C. Sullivan et al., *J. Biol. Chem.* **292**, 21149 (2017).
10. S. G. Hansen et al., *Science* **351**, 714 (2016).
11. S. G. Hansen et al., *Nat. Med.* **24**, 130 (2018).
12. J. Guan, S. J. Yang, F. Gonzalez, Y. Yin, N. Shastri, *J. Immunol.* **198**, 2017 (2017).
13. E. M. Doorduyn et al., *Front. Immunol.* **9**, 60 (2018).
14. M. Coppola et al., *Sci. Rep.* **6**, 37793 (2016).
15. Y. Bian et al., *PLOS Pathog.* **13**, e1006384 (2017).

ACKNOWLEDGMENTS

We thank T. van Hall for useful comments. Funded by ECHOR2020-TBVAC2020 (643381) and NIH (R21AI127133, R01AI141315).

10.1126/science.aay7079

ORGANIC CHEMISTRY

Photocatalytic deracemization fixes the mix

Photocatalysis converts racemic mixtures of *N*-aryl cyclic ureas into single enantiomers

By Alison E. Wendlandt

Very often, only one enantiomer of a chiral drug molecule is clinically active; the other optical isomer can produce unwanted or even catastrophic side effects. Thus, the efficient synthesis of single enantiomers is important in small-molecule drug synthesis, and resolution methods are frequently used to enrich a molecule's enantiomeric excess (ee) when racemic or scalemic mixtures (equal or unequal enantiomer mixtures, respectively) are obtained (1, 2). Classical and kinetic resolutions of mixtures can only reach a maximum 50% theoretical yield, and although dynamic kinetic resolution can achieve complete conversion to a single enantiomer, the resulting product is chemically modified (see the figure, top). Direct deracemization of chiral compounds reflects an ideal if rarely achieved approach to producing high-ee mixtures by resolution. On page 364 of this issue, Shin *et al.* (3) disclose a multicatalytic strategy for enriching racemic mixtures directly to a single enantiomer in a single step.

Deracemization methods are constrained by the principle of detailed balance, which dictates that any reverse chemical step must be in all ways microscopically identical to its corresponding forward step (4). Because (*R*) and (*S*) enantiomers are energetically identical, any elementary step or equilibrium process (catalyzed or not) that converts (*R*) into (*S*) must necessarily also convert (*S*) back into (*R*); the process will yield a racemate. Deracemization can be achieved, however, when

forward and reverse steps are designed to proceed by distinct mechanisms through a common prochiral intermediate (see the figure, bottom). Highly selective precedents are limited, although several notable approaches have relied on compartmentalization to separate incompatible oxidant and

reductant combinations (5, 6) or have used nonquenching redox pairs (7) to affect the desired deracemization in a single pot.

As Onsager noted in his seminal monograph (8), photochemical processes are among the handful of exceptions to the principle of detailed balance, in that forward and reverse steps may proceed along distinct electronic surfaces. A prominent example is the contra-thermodynamic photochemical isomerization of alkenes, wherein the selective absorption of light can lead to enrichment of the less thermodynamically stable constitutional isomer. This process constitutes the molecular basis for vision in vertebrates, and sensitized energy transfer has become a powerful synthetic strategy for obtaining (*Z*)-alkenes from (*E*)-alkenes through a triplet excited state (9). Extraordinary recent work by Bach and co-workers elegantly demonstrated this strategy for the catalytic deracemization of chiral alkenes (10).

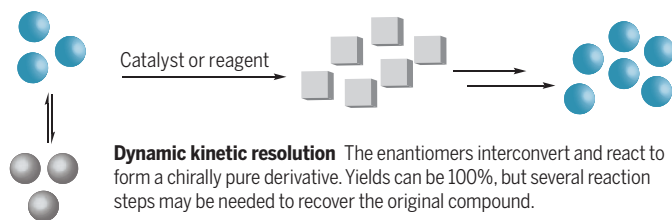
The method reported by Shin *et al.* takes a complementary photochemical approach to catalytic deracemization. Instead of energy transfer, the authors use an iridium photocatalyst to promote reversible electron transfer (ET) from the substrate. Enantioselective deprotonation (proton transfer, PT) of the resulting radical cation occurs in the presence of a chiral phosphate base to form a prochiral neutral radical intermediate with a net loss of a hydrogen atom from the substrate. This step functions as a photoinduced kinetic resolution, selectively ablating the stereochemical information of the faster-reacting (*S*)-substrate and enriching the once-racemic starting mixture in the (*R*)-substrate. The radical intermediate then reacts with a chiral thiol in an enantioselective hydrogen-atom

Deracemization versus resolution of enantiomers

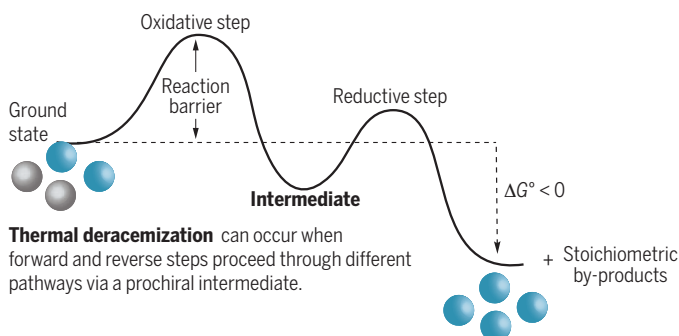
The fraction of a desired enantiomer (either *R* or *S*) from a mixture can be increased partially or indirectly through resolution methods, or interconverted completely through deracemization.



Kinetic (classical) resolution One enantiomer reacts faster to form a product Pr, selecting for the other enantiomer. Maximum yield is 50%.



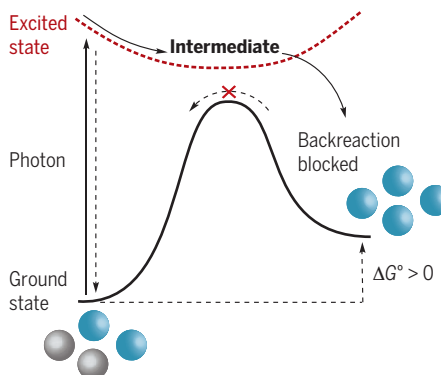
Dynamic kinetic resolution The enantiomers interconvert and react to form a chiral pure derivative. Yields can be 100%, but several reaction steps may be needed to recover the original compound.



Thermal deracemization can occur when forward and reverse steps proceed through different pathways via a prochiral intermediate.

Photochemical deracemization

In the route reported by Shin *et al.*, an iridium photocatalyst initiates electron transfer from *N*-aryl cyclic ureas. Subsequent reactions that include a chiral base convert the *S* to the *R* enantiomer.



transfer (HAT) step, this time kinetically favoring the formation of the (*R*)-substrate. Substoichiometric quantities of both the chiral base and chiral thiol can be used. The overall process consumes only photons, unlike thermal redox deracemizations, where stoichiometric oxidants and reductants are necessary to drive the reaction toward a single enantiomer.

The net result is a highly selective synthetic deracemization of *N*-aryl cyclic urea substrates. Yet remarkably, neither of the two individual steps (ET-PT and HAT) proceeds with high selectivity. Mathematically, the selectivity-determining step (SDS) is a composite of the two individual steps, and the observed enantiomeric ratio is the sum of the enantiomeric ratios of the two individual steps. The multiplicative form of the SDS results in extremely high overall enantiomeric enrichment through the synergistic action of these two chiral catalysts, even though each catalyst working individually is not particularly selective.

Although the scope of the reaction reported by Shin *et al.* is restricted to proof-of-concept substrates, undoubtedly exceptional cases such as this will become more

“...high overall enantiomeric enrichment through the synergistic action of these two chiral catalysts...”

common. In particular, the structural similarity of the current substrates to important classes of chiral ligands (for example, oxazolidinones) and to intermediates in enzymatic deracemization and stereoinversion pathways (for example, hydantoinase-decarbomoylase systems) hints toward the possibility of expanding this approach to the deracemization of amino acids and other important chiral molecules. ■

REFERENCES AND NOTES

1. J. M. Keith, J. F. Larrow, E. N. Jacobsen, *Adv. Synth. Catal.* **343**, 5 (2001).
2. H. Pellissier, *Adv. Synth. Catal.* **353**, 1613 (2011).
3. N. Y. Shin, J. M. Ryss, X. Zhang, S. J. Miller, R. R. Knowles, *Science* **366**, 364 (2019).
4. D. G. Blackmond, *Angew. Chem. Int. Ed.* **48**, 2648 (2009).
5. A. D. Lackner, A. V. Samant, F. D. Toste, *J. Am. Chem. Soc.* **135**, 14090 (2013).
6. M. Alexeeva, A. Enright, M. J. Dawson, M. Mahmoudian, N. J. Turner, *Angew. Chem. Int. Ed.* **41**, 3177 (2002).
7. Y. Ji, L. Shi, M.-W. Chen, G.-S. Feng, Y.-G. Zhou, *J. Am. Chem. Soc.* **137**, 10496 (2015).
8. L. Onsager, *Phys. Rev.* **37**, 405 (1931).
9. J. B. Metternich, R. Gilmour, *Synlett* **27**, 2541 (2016).
10. A. Hölzl-Hobmeier *et al.*, *Nature* **564**, 240 (2018).

10.1126/science.aay6919

PHOTOSYNTHESIS

Photosystem II, poised for O₂ formation

A consensus is emerging on the structure of the oxygen-evolving complex

By R. David Britt and David A. Marchiori

The vast majority of oxygen in Earth's atmosphere is generated by photosynthetic light reactions. This process also plays an integral role in producing energy for Earth's biosphere. Photosystem II (PSII) uses visible light to oxidize water and release O₂. Its oxygen-evolving complex (OEC) sequentially advances from its most reduced state (S₀), through four photon-driven oxidations, to its most oxidized state (S₄), which produces O₂. As described on page 334 of this issue, Suga *et al.* (1) used a powerful x-ray free-electron laser (XFEL) to examine in structural detail the OEC just prior to O=O bond formation, thereby adding to a growing body of work in this area. Such mechanistic details of the OEC's operations could help in the development of synthetic solar fuel reactions aimed at sustainable fuel production from water and sunlight.

The chemistry of the PSII water oxidation process has been the focus of studies coupling molecular biology and biochemistry to spectroscopic methods (2, 3). For many years, progress in the field was made without a structural basis for interpretation, but since the early 2000s the field has greatly benefited from steadily improved x-ray crystal structures of PSII nominally poised in the dark stable state (S₁) (4–6).

The S₁-state OEC is a Mn₄CaO₅ cluster, and the structural changes that occur in this cluster as it advances through the light-driven S-state transitions are a major question in the field. Coupled to these structural changes are questions concerning the S-state transitions in which the two substrate waters bind to the cluster and in what chemical form (e.g., H₂O, OH⁻, bridging, or terminal O²⁻). This is additionally complicated by the presence of an abundance of oxo bridges (a water-derived oxygen molecule that bridges two or more Mn/Ca ions) and water molecules within and proximal to the OEC. Which of these are the actual substrates that are oxidized to form O₂?

In one scenario, both substrate waters could bind to the OEC following O₂ release

in the S₄-to-S₀ transition. In this case they should be present as OEC ligands in the dark stable S₁ structure, where two waters (W1 and W2) are bound to a “dangler” Mn4 (7) and another two waters (W3 and W4) are bound to the Ca ion. This geometry suggests one often-discussed model for O=O bond formation, in which the dangler Mn4 (or another Mn of the cluster's core) is oxidized to a highly electrophilic Mn(V)-oxo species by the S₄ state. The Mn(V)-oxo could then be attacked by a nucleophilic water or hydroxide on the adjacent Ca ion (2, 5).

However, it is also plausible that one or both waters bind to the OEC later in the S-state cycle, arriving from some nearby solvent water site(s). Siegbahn's proposed model, based on density functional theory (DFT) energetics calculations, suggests that a water binds to an open coordination site at Mn1 in the S₂-to-S₃ transition (8), which upon further oxidation to S₄ forms a Mn(IV)O• species that creates the O=O bond through a radical coupling to the bridging O5 (see the figure). The conversion of a five-coordinate Mn(IV) to a six-coordinate Mn(IV) at the S₃ state, with a new water binding to Mn1 at S₃, has been supported by recent high-field electron paramagnetic resonance (EPR) spectroscopy studies (9, 10).

XFELs produce pulses of a few femtoseconds' duration with high enough photon energy and flux for use in electron diffraction of microscale protein crystals. This tool has been used to study the OEC's water oxidation mechanism. XFEL pulses are sufficiently short that the desired x-ray diffraction occurs before the microscale crystals are destroyed. The XFEL pulses can be synchronized with visible-light laser pulses to drive the S-state advancement. PSII is a particularly good target for this approach because short visible-light pulses drive the water-splitting reaction. This has allowed investigators to probe the structures of the transient S₂ and S₃ states, with several recent studies focused on the latter state, which is only 1 flash/electron transfer from the formation of O₂ (11–13).

There are many important aspects to these advanced S-state structures, but the question of substrate binding at the S₃ state is particularly crucial, as this is key to proposed

Department of Chemistry, University of California, Davis, CA 95616, USA. Email: rdbritt@ucdavis.edu

mechanisms for O=O bond formation. In 2016, a study reported a room-temperature structure of PSII following two light flashes, hence enriched in the S_3 state of the OEC (11). It reported no new S_3 -state electron density around Mn1, as one would expect for a newly bound water. This result directly conflicts with the proposed Siegbahn model and EPR studies. Shortly following this study, another group reported an S_3 structure with new electron density near Mn1 and O5 that was assigned to a newly bound oxygen (O6) (12). This was considered support for the S_3 -insertion substrate water model. However, the study showed close proximity (0.15 nm) between O5 and O6, consistent with O-O bond formation producing a peroxide at the S_3 state, which would in turn be associated with Mn reduction in the S_2 -to- S_3 transition. This is in conflict with EPR and x-ray spectroscopic data (3, 9, 14, 15). A more recent study reports a new oxygen “Ox” bound both to Mn1 and the Ca ion of the OEC (13), and with a 0.21 nm distance to O5, much longer than the reported 0.15 nm (12) and too long to represent the O-O bond of a peroxide.

Suga *et al.* have now cryotrapped serially flashed microcrystals of PSII with the goal of achieving more accurate interatomic distances, especially between O5 and O6. In the S_2 state, they observed an open cubane structure with a five-coordinate Mn1. At the S_3 state, a flip in the side chain of the monodentate carboxylate ligand Glu¹⁸⁹ (12) provides room for O6 to insert and bind to Mn1 and Ca. To address the crucial O5-O6 distance, Suga *et al.* calculated difference maps

as a function of modeled O5-O6 distance and found the smallest residual density at a distance of about 0.19 nm. Their full residual density analysis suggests an oxyl/oxo pairing for O5 and O6, which they consider evidence for an oxyl/oxo coupling mechanism in the O=O bond formation at the final S_4 state. Glu¹⁸⁹ appears to gate water insertion into the OEC from the “OI” water channel of the PSII reaction center.

Although there are still open questions as to the detailed assignments of substrate waters and the exact modes of water oxidation and O=O bond formation, the results from different studies are starting to converge, and the structural results appear to be closely in line with mechanistic proposals supported by computational chemistry and spectroscopy. The field anxiously awaits possible time-resolved structures detailing the O=O bond formation through the S_4 -to- S_0 transition. ■

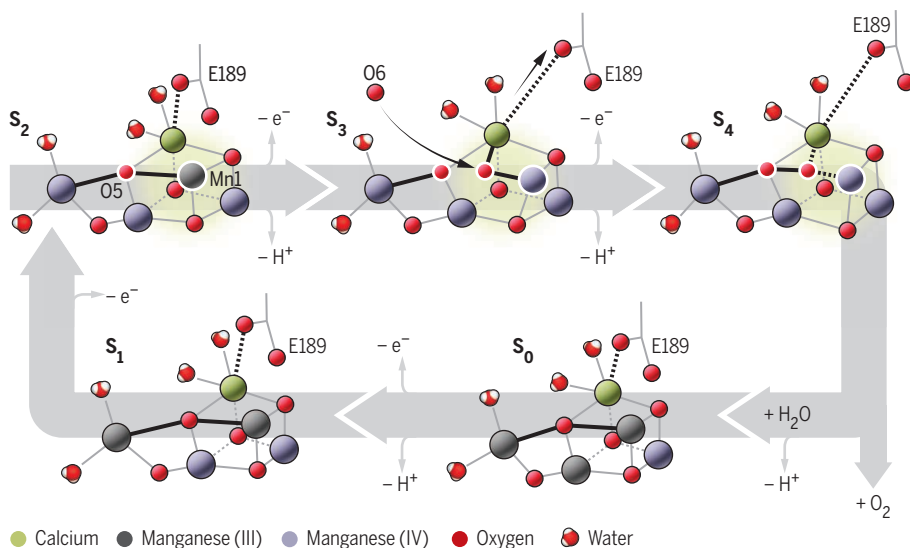
REFERENCES AND NOTES

1. M. Suga *et al.*, *Science* **366**, 334 (2019).
2. D. J. Vinyard, G. W. Brudvig, *Annu. Rev. Phys. Chem.* **68**, 101 (2017).
3. D. A. Pantazis, *ACS Catal.* **8**, 9477 (2018).
4. A. Zouni *et al.*, *Nature* **409**, 739 (2001).
5. K. N. Ferreira *et al.*, *Science* **303**, 1831 (2004).
6. Y. Umena *et al.*, *Nature* **473**, 55 (2011).
7. J. M. Peloquin *et al.*, *J. Am. Chem. Soc.* **122**, 10926 (2000).
8. P. E. M. Siegbahn, *Biochim. Biophys. Acta Bioenerg.* **1827**, 1003 (2013).
9. N. Cox *et al.*, *Science* **345**, 804 (2014).
10. M. Chrysinia *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16841 (2019).
11. I. D. Young *et al.*, *Nature* **540**, 453 (2016).
12. M. Suga *et al.*, *Nature* **543**, 131 (2017).
13. J. Kern *et al.*, *Nature* **563**, 421 (2018).
14. M. Haumann *et al.*, *Science* **310**, 1019 (2005).
15. N. Schuth *et al.*, *Inorg. Chem.* **57**, 10424 (2018).

10.1126/science.aaz4522

The S states in the oxygen-evolution reaction

The oxygen-evolving complex is photo-oxidized through a series of S states to produce molecular oxygen from water. In the final steps before O=O bond formation, a new oxygen, O6, binds to the vacant site at Mn1. After a final photo-oxidation event, O5 and O6 appear poised to form an O=O bond, releasing molecular oxygen, reducing the cluster, and beginning the catalytic cycle anew. Glutamic acid at position 189 is noted as E189.



NEUROSCIENCE

Spikes in the sleeping brain

Memory is replayed and consolidated under low background noise during deep sleep

By Yuji Ikegaya^{1,2} and Nobuyoshi Matsumoto¹

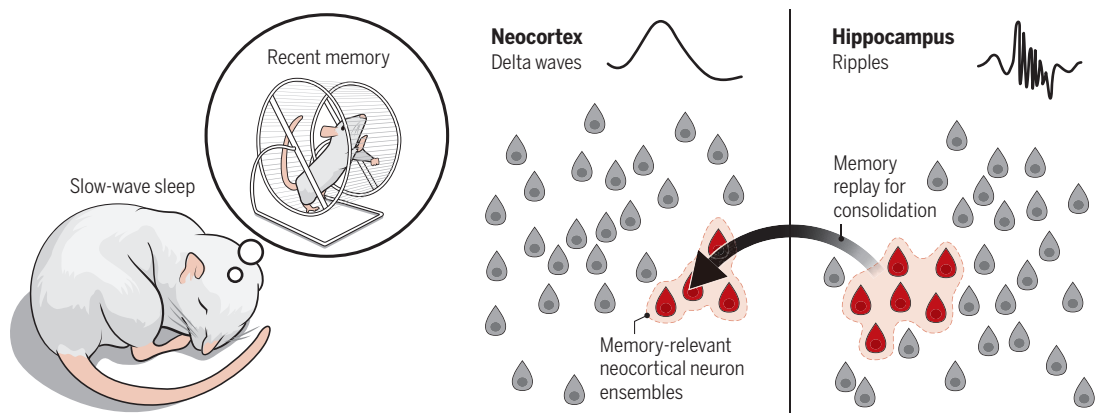
Memories of recent experiences are transferred and fixed in long-term storage in the neocortex during sleep. A key aspect of this process is the occurrence of ripples, which are high-frequency oscillations in neuronal activity in the hippocampus (1, 2). Ensembles of hippocampal neurons that have been activated during an experience emit synchronized neuronal activity (spikes) during ripples in subsequent sleep (3). These “replayed” memories propagate to the neocortex during slow-wave states (4), which are characterized by alternating “bright” states with active neurons and “dark” states, called delta waves, without active neurons (5). Researchers reasoned that hippocampal-cortical dialogue occurs during the bright states and that delta waves represent intermittent neocortical “sleep” to recover from synaptic fatigue (6) or increased potassium conductance (7) and have no active function. On page 377 of this issue, Todorova and Zugaro (8) challenge this view, demonstrating that a small number of neurons are reactivated during delta waves.

Todorova and Zugaro recorded activity, both spikes from multiple neurons and local field potentials, simultaneously in the rat dorsal hippocampus and the medial prefrontal cortex, which is believed to store long-term memories. They reconfirmed that the prefrontal cortex becomes inactive during each delta-wave epoch but also discovered that a few neurons remain active and occasionally emit synchronous spikes during delta waves (which they called “delta spikes”) (see the figure). The neurons recorded in this study constitute only a small portion of the total cells in the prefrontal cortex. Thus, it is important to ask how frequently delta spikes occur. The local field

¹Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo 113-0033, Japan. ²Center for Information and Neural Networks, National Institute of Information and Communications Technology, Suita City, Osaka 565-0871, Japan. Email: yuji@ikegaya.jp

Hippocampal-cortical dialogues during sleep

The hippocampus communicates with the neocortex and activates experience-associated neocortical neuron ensembles during delta waves, low-cerebral activity states that emerge intermittently during slow-wave sleep. This selective activity focuses memory consolidation.



potential waveforms, which reflect the synchronous activity of a group of neurons, do not differ depending on whether delta waves are accompanied by delta spikes in the recorded neurons. Moreover, the probability of detecting delta spikes is proportional to the number of simultaneously recorded neurons. Extrapolating these observations suggests that delta spikes accompany virtually all delta waves.

Delta spikes are not shots in the dark or remnants of ripples that accidentally fail to disappear during delta waves. Rather, delta spikes occur in spatiotemporally designed, specifically recursive configurations and seem to be actively generated in the brain. Todorova and Zugaro annotated delta spikes as behavioral correlates. As rats were trained in a spatial memory task, behaviorally relevant neurons became preferentially activated in delta waves during the following sleep period. The neuronal replay in the prefrontal cortex appears to be instructed by hippocampal ripples, because the set of prefrontal neurons that emit a delta spike in a given delta wave can be statistically predicted from the set of hippocampal neurons activated in the ripple immediately preceding the delta wave. The spike correlations of cell pairs between the hippocampus and the medial prefrontal cortex remain high for hundreds of milliseconds, indicating that hippocampal-cortical dialogue occurs within a short time frame, during which the neocortex serves as a temporary reservoir of hippocampal information. As a result, memory-relevant delta spikes are fired mainly in the early phase of each delta-wave event.

Delta spikes consist mainly of behaviorally relevant neurons, whereas behaviorally irrelevant neurons are more likely silenced during delta waves. Reducing irrelevant

information improves the signal-to-noise ratio of neuronal information and thereby helps isolate cortical computations to avoid deleterious cross-talk. That is, the neocortex quiets down during delta waves and focuses attention on the content to be memorized, consistent with the authors' previous prediction that delta waves "isolate target synapses from competing inputs, allowing selective reorganization of the network" (9). This leads to the possibility that facilitating delta spikes in ripple-associated neurons could enhance memory consolidation. Consistent with this idea, task performance in mice is enhanced when artificial delta waves containing ripple-coupled delta spikes are repeatedly induced using brief single-pulse electrical stimulation of deep cortical layers during sleep after an object-discrimination task. This suggests that delta spikes mediate memory consolidation.

"...task performance in mice is enhanced when artificial delta waves...are repeatedly induced..."

How do delta spikes contribute to memory consolidation? It is unknown what delta waves give rise to in prefrontal neuronal circuits. Delta spikes are transient and thus are unlikely to work as persistent memories by themselves. So, what form of long-lasting plasticity is induced by delta spikes? Moreover, activity patterns in individual task-associated neocortical neurons change over a period of days after learning (10). Do delta spikes also have an instructive role in the dynamic reorganization of neuronal identities? Another line of evidence suggests that delta waves are involved in forgetting rather than memory consolidation (11). Thus, drifting delta spikes could contribute to deactivation of memory traces.

Another question is how the dorsal hippocampus and the medial prefrontal cortex communicate during slow-wave sleep. Direct synaptic connections are unlikely

to exist between these brain regions. On the basis of anatomical evidence of axonal projections, candidates for the relay station include the ventral hippocampus (12), a neocortical area via the entorhinal cortex (13), and the retrosplenial cortex via the subiculum (14). Recording or ablation of neuronal activity in these brain regions is necessary to answer this question.

Delta waves are a hallmark of the entire neocortex during sleep, and delta spikes may occur in cortical regions other than the medial prefrontal cortex. Do delta spikes have a role in processes other than memory consolidation? Todorova and Zugaro have successfully linked their unexpected discovery to the current knowledge of memory and sleep, but future investigations are still required to uncover the precise mechanisms and functions of delta spikes. ■

REFERENCES AND NOTES

1. G. Girardeau, K. Benchenane, S. I. Wiener, G. Buzsáki, M. B. Zugaro, *Nat. Neurosci.* **12**, 1222 (2009).
2. A. Fernández-Ruiz et al., *Science* **364**, 1082 (2019).
3. G. Buzsáki, *Neuroscience* **31**, 551 (1989).
4. N. K. Logothetis et al., *Nature* **491**, 547 (2012).
5. F. Siclari, G. Tononi, *Curr. Opin. Neurobiol.* **44**, 222 (2017).
6. D. Contreras, I. Timofeev, M. Steriade, *J. Physiol.* **494**, 251 (1996).
7. M. V. Sanchez-Vives, D. A. McCormick, *Nat. Neurosci.* **3**, 1027 (2000).
8. R. Todorova, M. Zugaro, *Science* **366**, 377 (2019).
9. N. Maingret, G. Girardeau, R. Todorova, M. Goutierre, M. Zugaro, *Nat. Neurosci.* **19**, 959 (2016).
10. L. N. Driscoll, N. L. Pettit, M. Minderer, S. N. Chetthi, C. D. Harvey, *Cell* **170**, 986 (2017).
11. J. Kim, T. Gulati, K. Ganguly, *Cell* **179**, 514 (2019).
12. A. M. Thierry, Y. Gioanni, E. Dégénétais, J. Glowinski, *Hippocampus* **10**, 411 (2000).
13. R. Insausti, M. T. Herrero, M. P. Witter, *Hippocampus* **7**, 146 (1997).
14. M. P. Witter, R. H. Ostendorf, H. J. Groenewegen, *Eur. J. Neurosci.* **2**, 718 (1990).

ACKNOWLEDGMENTS

This work was supported by Japan Science and Technology Agency (JST) ERATO JPMJER1801 and Japan Society for the Promotion of Science (JSPS) KAKENHI 18H05525.



ECOLOGY

Rapid reorganization of global biodiversity

Marine systems outpace terrestrial habitats in biodiversity erosion

By **Britas Klemens Eriksson**¹ and **Helmut Hillebrand**^{2,3}

Twenty-five years of research on the relationship between biodiversity and ecosystem function have revealed that biodiversity drives fundamental ecosystem processes and regulates their temporal and spatial stability (1, 2). Despite clear signs that human efforts have failed to halt global biodiversity loss (3, 4), it has been difficult to identify corresponding signs of global-loss trends in the context of local ecosystems (5–9). On page 339 of this issue, Blowes *et al.* (10) report their analysis of local biodiversity changes using a large dataset of >50,000 biodiversity time series from 239 studies. Each time series represents a record of species composition at a selected site over time, with sites representing all major ecosystem types and climatic zones. The authors demonstrate that the identities of species and their abundances are being rapidly reorganized.

Although the proportion of reported global extinctions remains small compared with previous mass extinctions, the current decline in the abundance of myriad

species and the elevated danger of their extinction have raised scientific concern over the ill effects of biodiversity loss on human health and well-being (3). The controversy over global versus local changes in biodiversity arose from comprehensive analyses of trends in species numbers worldwide; the data revealed that local communities show strong species turnover but no systematic decrease in the total number of species—a metric referred to as species richness. In fact, most habitats undergo replacement of some or many species by other distinct species with no general shift in the total number of species (5, 6). Even coastal time series show a systematic increase in average species richness over time, and declines are observed only in locations with strong human pressures (7).

Together, these studies (5–7) initiated a debate on why the average number of species in local communities has not (yet) been negatively affected by the global erosion of biodiversity (5–9, 11). A key lesson learned from this debate is that biodiversity change represents much more than a change in the average number of species.

Species richness in a local habitat reflects the net difference between the num-

ber of species immigrating into a habitat and the number becoming extinct over time, and it is not a sensitive measure of biodiversity change (6, 7), which also includes shifts in species composition and abundances of individual species types (see the figure). A recent supporting study found that a complete exchange of species composition over time without any change in the number of species (that is, species richness) commonly occurs in nature (11).

The same study challenged the expectation that changes in ecosystem conditions (for better or worse) will be reflected by simple trends in the number of species; in fact, this hypothesis invokes an equilibrium view of biodiversity that might rarely be achieved. A habitat undergoing environmental change can become suitable for new species, but their impact on resident species will take time to manifest (see the figure). Thus, native species are likely to survive for some time before they are excluded (for example, through competition). Ecological systems exposed to continuous environmental change can, therefore, be expected to show a systematic increase in the number of species if the rate of immigration is faster than the rate of extinction (so-called extinction debt) (12).

¹Marine Ecology, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, Netherlands. ²Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Wilhelmshaven, Germany. ³Helmholtz Institute for Functional Marine Biodiversity Oldenburg and Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany. Email: b.d.h.k.eriksson@rug.nl; helmut.hillebrand@uni-oldenburg.de

The Pacific oyster is native to Pacific Asia but has spread through accidental introductions across the world. Today, it is found on shores in Australia, Europe, New Zealand, and North America.

The need to monitor biodiversity changes by addressing not only species identities but also their abundances is intensified by a traditional problem in the ecological and environmental sciences: Most data from relevant monitoring programs are locked up in governmental programs and institutions and are not openly accessible by scientists and the public. To address this serious deficiency, Dornelas and colleagues developed a community-led open-source database of biodiversity time series called BioTIME, with the aim of promoting robust analyses of changes in global biodiversity (13).

Using BioTIME, Blowes *et al.* present a dynamic perspective of biodiver-

sity changes that will shift the discussion on this topic. They show that the rate of change is highly context-dependent, but they also demonstrate increasing instability of biodiversity, which raises concerns in an era of global change. In particular habitats, the number of species either increased or decreased over time, but when assessing all habitats simultaneously, the authors observed an increase in the average number of species across the globe. The most notable finding by Blowes *et al.* was that local species assemblages were being rapidly reorganized over time, with nearly a third of all species being replaced by new species every decade; these changes in biodiversity were stronger in marine systems than in terrestrial ones. Their study thus highlights that the global biodiversity crisis, at least for now, is not primarily about decline but, rather, about large-scale reorganization.

These findings raise questions about species homogenization and adaptability (14), as well as an expectation of precarious ecosystem services in the future, when a large reorganization of local species assemblages affects the temporal stability of ecosystem functions (see the figure). Human-mediated introduction of new species might increase local species richness by elevating immigration rates, but at the same time it promotes biological homogenization—the process by which the composition of species in an ecosystem becomes more uniform. Homogeneity in local species leads to erosion of biodiversity over larger regions (15) and can limit the ability of species communities to adapt to future environmental changes if immigration rates decline (14). Thus, local increases in species richness might even correlate with regional declines in ecosystem resilience over time. Furthermore, by showing high temporal turnover of species composition for all possible trends in species richness, Blowes *et al.* confirm that such trends are insufficient to describe the consequences of global change for ecological communities.

The study of Blowes *et al.* also has two implications for ecology as a scientific discipline: It brings together authors who originally had highly divergent views on local biodiversity change, and it demonstrates the importance of open access to biodiversity data of sufficient quality. Only the high quality of BioTIME data enabled Blowes *et al.* to clearly identify geographical regions that experienced more or less change in biodiversity than expected, which will affect the setting of conservation priorities. With open-source access to international databases, the scientific community can more precisely monitor the biodiversity crisis and develop local solutions. ■

REFERENCES AND NOTES

1. F. Pennekamp *et al.*, *Nature* **563**, 109 (2018).
2. J. E. Duffy, C. M. Godwin, B. J. Cardinale, *Nature* **549**, 261 (2017).
3. IPBES, "Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services" (IPBES Secretariat, 2019).
4. D. P. Tittensor *et al.*, *Science* **346**, 241 (2014).
5. M. Vellend *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19456 (2013).
6. M. Dornelas *et al.*, *Science* **344**, 296 (2014).
7. R. Elahi *et al.*, *Curr. Biol.* **25**, 1938 (2015).
8. A. Gonzalez *et al.*, *Ecology* **97**, 1949 (2016).
9. M. Vellend *et al.*, *Ecology* **98**, 583 (2017).
10. S. A. Blowes *et al.*, *Science* **366**, 339 (2019).
11. H. Hillebrand *et al.*, *J. Appl. Ecol.* **55**, 169 (2018).
12. S. T. Jackson, D. F. Sax, *Trends Ecol. Evol.* **25**, 153 (2010).
13. M. Dornelas *et al.*, *Glob. Ecol. Biogeogr.* **27**, 760 (2018).
14. D. Hodapp *et al.*, *Ecol. Lett.* **21**, 1364 (2018).
15. T. Finderup Nielsen, K. Sand-Jensen, M. Dornelas, H. H. Bruun, *Ecol. Lett.* **22**, 1650 (2019).

Interplay of local and regional diversity alters environment

Three biodiversity scenarios occur at different times in history: equilibrium scenario (past; no global change in biodiversity), current scenario (moderate global change), and future scenario (both regional and local richness decrease over time). Each symbol and color depicts a distinct species.

Regional biodiversity (past)

Immigration
Extinction
Emigration

Local biodiversity

Environmental filter

Colonization

Equilibrium scenario

A balance among species immigration, emigration, and extinction events regulates regional biodiversity. Members of the regional species pool pass through an environmental filter and colonize local species assemblages.

Regional biodiversity (current)

Biological homogenization

Local biodiversity

Current scenario

Environmental factors cause the growth of some species to be favored over that of others. Because the environmental impact takes time to manifest, threatened species likely will survive for a while before becoming extinct. This leads to an increase in local species richness despite regional erosion of biodiversity.

Regional biodiversity (future)

Biological homogenization

Local biodiversity

Future scenario

With prolonged human impact, environmental change intensifies, and species resilience erodes. Populations of new, better-adapted species grow, and regional extinctions intensify. If a local region is isolated or the global pool lacks a richness of adapted species, immigration cannot compensate for biodiversity loss.

POLICY FORUM

SYNTHETIC BIOLOGY

Technological challenges and milestones for writing genomes

Synthetic genomics requires improved technologies

By Nili Ostrov¹, Jacob Beal², Tom Ellis³, D. Benjamin Gordon⁴, Bogumil J. Karas⁵, Henry H. Lee¹, Scott C. Lenaghan⁶, Jeffery A. Schloss⁷, Giovanni Stracquadanio⁸, Axel Trefzer⁹, Joel S. Bader¹⁰, George M. Church¹¹, Cintia M. Coelho¹², J. William Efcavitch¹³, Marc Güell¹⁴, Leslie A. Mitchell¹⁵, Alec A. K. Nielsen¹⁶, Bill Peck¹⁷, Alexander C. Smith¹⁸, C. Neal Stewart Jr.¹⁹, Hille Tekotte²⁰

Engineering biology with recombinant DNA, broadly called synthetic biology, has progressed tremendously in the last decade, owing to continued industrialization of DNA synthesis, discovery and development of molecular tools and organisms, and increasingly sophisticated modeling and analytical tools. However, we have yet to understand the full potential of engineering biology because of our inability to write and test whole genomes, which we call synthetic genomics. Substantial improvements are needed to reduce the cost and increase the speed and reliability of genetic tools. Here, we identify emerging technologies and improvements to existing methods that will be needed in four major areas to advance synthetic genomics within the next 10 years: genome design, DNA synthesis, genome editing, and chromosome construction (see table). Similar to other large-scale projects for responsible advancement of innovative technologies, such as the Human Genome Project, an international, cross-disciplinary effort consisting of public and private entities will likely yield maximal return on investment and open new avenues of research and biotechnology.

The ability to design and write genomes of living cells provides distinctive opportunities to tackle problems that are intractable with current technologies. These transformative technologies could have widespread scientific, social, and economic impacts; thus, their development and adoption require proactive identification of potential pitfalls through ongoing public discussion. Accordingly, the outlook presented here is part of a broader effort by the international Genome Project-write (GP-write) consortium (1) to encourage inclusive conversation among scientists,

lawyers, ethicists, educators, environmentalists, other experts, and stakeholders, as well as the general public, and ensure responsible, safe, and coordinated implementation of these new technologies.

Synthetic genomics is a relatively new field, and the majority of writing technologies (with the exception of commercial DNA synthesis) are still developed by academic research laboratories. Thus, unlike the reasonably predictable progress of engineering in more established industries, such as semiconductors, prediction of timelines and costs in these nascent fields remains highly speculative (see table).

GENOME DESIGN

Genome design aims to encode higher-level design criteria into DNA sequences at the chromosome scale. This will require computer-aided design (CAD) technologies to (i) reliably produce desired phenotypes; (ii) maximize the impact of the design, both in terms of experimental feedback and technical feasibility; and (iii) facilitate collaboration by employing standards for handling and exchanging design information. Current synthetic genomic CAD software, such as used for Sc2.0 (2), uses automation and collaboration tools for scaling DNA design from plasmids to entire chromosomes. Currently, estimation of functional effects of edits (e.g., gene deletions) is left to human experts. Looking forward, these tools will need to accurately predict the viability and phenotype of a cell from its genome design.

Although simple models are sufficient to handle silent edits (e.g., synonymous editing of coding sequences for watermarking), models of increasing complexity and precision will be necessary to predict how changes to genome sequence affect gene regulation and protein function. Although comprehensive mechanistic models of higher eukaryotes are likely decades away, machine learning approaches could improve pheno-

type prediction by leveraging the wealth of high-dimensional, high-throughput systems biology data in public databases and generated from genome writing projects, similar to how such techniques have been used to predict protein structures.

Leveraging these models, there is a need for experimental design tools to minimize the number of expensive iterations required to obtain successful genome designs. For instance, redesign of the relatively small *Mycoplasma* genome required four iterations, as well as genomic screens to identify essential genes (3). Larger projects will require many more intermediate stages of construction and supporting experiments. To provide the most valuable feedback and leverage the resulting data in subsequent designs, new algorithms are required to automate design of experiments and selection of appropriate engineering technologies for their implementation (e.g., “write versus edit”).

A related, unaddressed need is to ensure compatibility of designed DNA with constraints from downstream synthesis, assembly, delivery, and analytical stages (e.g., software-guided parsing of chromosome sequence into synthesis-compatible fragments, or introduction of designated sequences to facilitate assembly and delivery). Tools will need to be adaptable to anticipate advances in writing technologies.

All of these efforts rely on and generate large datasets, which require stewardship to streamline model definition and facilitate sharing of results. Two major barriers for integrating biological information are data incompatibility and a lack of sufficiently descriptive metadata. We encourage researchers to use widely adopted data-exchange formats, e.g., GenBank and general feature format, and to continue to establish and adhere to standards for experimental metadata, e.g., the Synthetic Biology Open Language (SBOL).

Funding agencies and industrial stakeholders should prioritize support for long-term development of software and standard data formats, including collaboration, visualization, and quality control capabilities, similar to initiatives already in place for genomics, such as the Wellcome Trust Open Research Fund and the Chan Zuckerberg BioHub. Synthetic genomics software will not only enhance our ability to plan and execute large-scale genetic projects, but will likely also lead to fundamental methodological advances to move from correlation to causation of genotype-phenotype relationships.

DNA SYNTHESIS

Genome writing projects depend on large numbers of long (>5000 base pairs (bp)) and precise synthetic DNA constructs (4, 5). How-

Author affiliations are listed in the supplementary materials.
Email: schlossjeff500@gmail.com

Key challenges and milestones for synthetic genomes

KEY TECHNOLOGY DEVELOPMENT TARGET	EXAMPLE OF DESIRED MILESTONES	ESTIMATED TIME (YEARS)
Genome design		
Develop tools for genome-scale design, visualization, and quality control.	Design a virus-proof mammalian chromosome.	3
Integrate structural information (2D and 3D) into genome design software	Predict the conformation of a synthetic yeast chromosome.	5
Develop sequence-to-phenotype whole-cell modeling.	Optimize metabolic profile, accurate to within twofold, for 100 key gene products of a synthetic virus-proof chromosome.	10
DNA synthesis		
Increase coupling efficiency for oligonucleotide synthesis.	Synthesize high-fidelity oligonucleotides longer than 500 nucleotides.	3
Increase efficiency of in vitro DNA assembly for fragments >20 kb.	Assemble 20 kb with >50% yield.	4
Develop methods for synthesis of difficult sequences, including homopolymers, high-GC content, and secondary structure.	Synthesize a centromere.	5
Develop enzymatic methods for direct synthesis of multikilobase DNA fragments.	Synthesize a 10-kb fragment (without assembly).	7
Decrease cost of DNA synthesis by 1000-fold.	Synthesize and assemble DNA for one haploid human genome (i.e., 3.2×10^9 bases) for \$1000.	10
Genome editing		
Expand multiplexity and precision of DNA editing.	Simultaneously edit 1000 different targets in a single bacterial, mammalian, or plant cell with 1 off-target hit per 10,000 genomes.	2
Increase efficiency of homologous-directed repair (HDR)-mediated editing in mammalian and plant cells.	Perform HDR-mediated editing in nondividing mammalian cells at >90% efficiency.	3
Develop editing enzymes for precise substitution of any nucleotide at any desired genomic locus, with increased efficiency.	Perform allele editing of human cells at sites lacking PAM sequence, with >95% efficiency.	5
Chromosome construction		
Develop methods for temporal and spatial control of single chromosomes, such as chromatin state.	Engineer segregating, stable human artificial chromosomes (HAC).	2
Develop specialized host cells with high efficiency for DNA assembly, particularly for difficult-to-assemble sequences.	Establish in vivo chromosome assembly methods in the host <i>Streptomyces coelicolor</i> (72% GC content).	5
Develop efficient, inexpensive methods for routine and automated delivery of entire chromosomes into cells.	Demonstrate routine, device-based chromosome delivery in mammalian cells by cell fusion.	3
Develop methods for assembly and testing of Mb-size chromosomes.	Assemble a synthetic recoded human chromosome 21 from DNA fragments.	10

ever, chemical synthesis of DNA remains limited to production of short oligonucleotides (oligos), commonly 200 bp long. Although oligos have driven major advances in recombinant DNA technologies, larger DNA constructs require assembly of multiple oligos, a process that is laborious and lossy. Routine production of long, precise fragments of synthetic DNA would be desirable.

Although DNA has become more available through commercial vendors in recent years as a result of industrialization by parallelization and miniaturization, there has been little improvement to the underlying phosphoramidite chemistry, which limits DNA length, production speed, and cost. Accordingly, construction of whole chromosomes remains expensive and time-consuming. For example,

array synthesis of oligos costs approximately \$0.0005 per nucleotide, yielding an estimate of \$1.5 million for synthesizing 3 gigabases of DNA—the size of a human genome. Radical new approaches to DNA assembly, purification, and synthesis processes are thus required to achieve substantial advancement on cost and ease.

Innovations to minimize or eliminate the need for assembly, error correction, and cloning of DNA fragments assembled from oligos could boost productivity of current DNA synthesis infrastructure. To increase the yield of perfect sequences, currently ranging from 5 to 60% (6), hybridization and error correction can benefit from the engineering of high-fidelity polymerases and ligases. These advancements, largely

driven by industry, will decrease operating costs and production time. Cloning efficiency can also be enhanced by harnessing hosts with rapid division and/or high recombination rates, or obviated by using cell-free cloning and artificial cells. These technologies require fundamental research before they reach commercial readiness.

New technologies capable of synthesizing high-quality long DNA fragments would fundamentally alter chromosome-scale engineering. Recently, rapid production of short sequence-defined single-stranded DNA (ssDNA) has been reported by using the template-independent DNA polymerase TdT (terminal deoxynucleotidyl transferase) (7, 8). TdT offers the potential for directly synthesizing multikilobase sequences with increased polymerization rate and higher coupling efficiencies. To compete with existing phosphoramidite chemistry, enzymatic synthesis approaches should be further developed to address complex sequences and achieve precise, high-quality DNA in an automated and cost-effective fashion. These efforts, ripe for startup company attention, will benefit from continued investment in fundamental research to elucidate molecular mechanisms of enzymatic terminal transferase reactions.

To support the scale and quality of DNA required for genome-scale projects, enhancement of electromechanical systems as well as innovative biological tools are needed. Continued increase in throughput may be achieved by further parallelization and miniaturization—for example, by semiconductor fabrication or droplet-based techniques. Increases in DNA quality and production speed will be influenced by use of biological tools such as enzymes and organisms.

GENOME EDITING

Powerful new DNA editing tools have lowered the technical barriers for performing highly precise genetic and epigenetic modifications. Multiplexed editing of an intact genome could sharply decrease the time and labor required to generate a large number of modifications and, in some instances, circumvent the need for de novo synthesis and chromosome assembly.

Yet, despite considerable success with programmable nucleases such as Cas9, TALEN, and ZFN in multiple cell types with exquisite temporal and tissue-specific control, genome-scale editing remains limited. A localized, nuclease-induced double-strand break can be used to increase editing efficiency at each locus, but multiple simultaneous breaks often cause cellular toxicity. To avert toxicity, “base editor” enzymes were engineered in which the nuclease is replaced by enzymatic base modification (9), achieving simultaneous editing of over 13,000 Alu

repetitive elements in human cells by using a small number of guides (10). Other engineered Cas9 tools are used for repression, activation, or targeted insertion of DNA. However, a major bottleneck for multiplex genome-wide editing remains the delivery of guide RNAs (gRNAs), as multiple unique genome changes necessitate the presence of multiple unique gRNAs in the same cell. We anticipate that this barrier, as well as off-target mutagenesis and constraints caused by sequence specificity of editing enzymes [e.g., protospacer adjacent motif (PAM) sequence requirement], will be overcome to enable routine multiplex editing.

Genome-scale editing can also be accomplished by oligo recombineering (11), which relies on homologous recombination (HR) and has reduced in vivo toxicity. However, this technique is currently limited to a handful of organisms in which high-efficiency HR can be catalyzed by a recombinase that uses a donor ssDNA for targeting. To enable recombineering to edit plants and mammalian cells, new recombinase enzymes must be discovered or designed. It may also be necessary to map and modulate an organism's repair pathways to improve HR.

Additionally, comprehensive suites of molecular tools should be generated to accelerate testing and optimization of genome editing. For example, programmable TALEN or ZFN nucleases can be generated for targeting all UAG stop codons in human cells. Similarly, CRISPR-Cas9 guide libraries targeting all PAMs can be used to explore multiplexed, allele-specific targeting in plant, human, or fungal cells. Efforts to generate these genome-wide resources will provide experimental evidence of "accessibility maps" that reflect editing efficiency variability across genomic targets. Such data will optimize the choice of target sequences, inform predictive computational models, and deepen our knowledge of chromosome structure, folding, and repair.

CHROMOSOME CONSTRUCTION

The most critical hurdle facing synthetic genomics is the assembly and introduction of synthetic chromosomes into host cells. How does one stitch together all the DNA pieces required to construct a fully functional chromosome? Once constructed, how can we control chromosome localization and architecture to ensure cell viability? How do we replace all chromosome copies in polyploid organisms? As the genomes of most free-living organisms are larger than 2 Mb, methods for routine manipulation of large DNA fragments are critically needed.

Despite recent improvements in DNA synthesis and in vitro cloning, such methods are not efficient for construction of en-

tire chromosomes. Higher-order assembly of chromosomes at least 1 Mb in length can be performed by in vivo HR in the yeast *Saccharomyces cerevisiae*—a robust technique used in all synthetic chromosomes reported to date, including viral, bacterial, yeast, and algal chromosomes, as well as fragments of mice and human genomes (12, 13).

The efficiency of DNA assembly in *S. cerevisiae* has not been found in other genetically tractable organisms. To expand the toolkit for writing specialized synthetic chromosomes that are difficult to assemble in *S. cerevisiae*, new HR-proficient cloning organisms should be developed that tolerate high GC, direct repeats, and desired posttranscriptional modifications. Organisms compatible with extreme environments such as desert, deep ocean, or space travel may provide new routes for DNA assembly, such as found in the polyextremophile *Deinococcus radiodurans*.

Once a chromosome is constructed, delivery and manipulation become the primary engineering bottleneck in most desired hosts. To deliver megabase-scale constructs, robust, high-throughput DNA transformation methods must be developed in a variety of organisms spanning genera. For example, breakthroughs in DNA delivery can revolutionize plant engineering, which is currently hindered by species-specific, labor-intensive transformation methods and limited by traditionally conservative funding. High-risk, high-reward funding to support modernization of plant research, such as development of tissue-culture-independent DNA delivery methods, is pertinent for synthetic biology enabled improvements of agricultural organisms. Automation of specialized methods for chromosome transfer between yeast, bacteria, plants, and mammalian cells (such as cell fusion, genome transplantation, or microinjection) requires multidisciplinary funding opportunities aimed at bridging microfluidics with traditional cell and molecular biology work. It is essential that early proof-of-concept efforts be supported at government and foundation levels.

Many cellular forces that shape genome structure and function remain largely unknown. Fundamental studies are needed to elucidate mechanisms by which sequence and epigenetic regulation guide inter- and intrachromosomal interactions and determine genome architecture. Emerging technologies for programmable modifications of chromatin, such as insulators guiding chromatin remodeling, safe harbor sites for DNA insertion, and orthogonal recombinase enzymes, will be necessary for developing gene therapies (14). Better understanding of organelle genomes (plastid, mitochondrion), which remain extremely difficult to engineer, would offer new routes

for stable maintenance and incorporation of artificial chromosomes.

A final challenge when introducing synthetic constructs is to quickly determine whether they perform as desired in the destination cell. Whole-genome DNA and RNA sequencing will serve as a first-pass verification of chromosome integrity. Tailored cell lines with phenotypic reporters may be developed for assessing the performance of large synthetic constructs. Reliable organoid models and a clear understanding of regulation and expression changes that drive organismal development will be key to extrapolating results from single cells to the design of functional chromosomes for multicellular organisms.

GLOBAL AND MULTIDISCIPLINARY

New technologies may come from efforts in synthetic biochemistry, such as programmable synthetic protocells, from progress at the interface of hardware and wetware such as solid-phase DNA assembly platforms, or from findings in basic bioscience research—for example, by uncovering valuable new enzymes or delivery systems. Innovation will be driven by government grants and genomic and cancer institutes, with a growing role for BioFoundries, emerging hubs for automation of bioengineering. A highly interdisciplinary, multinational effort from government and private sectors will help achieve and disseminate these advances to make an impact in biomedical, pharmaceutical, agricultural, and chemical industries. ■

REFERENCES AND NOTES

1. J. D. Boeke *et al.*, *Science* **353**, 126 (2016).
2. S. M. Richardson *et al.*, *Science* **355**, 1040 (2017).
3. C. A. Hutchison 3rd *et al.*, *Science* **351**, aad6253 (2016).
4. J. Fredens *et al.*, *Nature* **569**, 514 (2019).
5. N. Ostrov *et al.*, *Science* **353**, 819 (2016).
6. N. B. Lubock, D. Zhang, A. M. Sidore, G. M. Church, S. Kosuri, *Nucleic Acids Res.* **45**, 9206 (2017).
7. S. Palluk *et al.*, *Nat. Biotechnol.* **36**, 645 (2018).
8. H. H. Lee *et al.*, *Nat. Commun.* **10**, 2383 (2019).
9. H. A. Rees, D. R. Liu, *Nat. Rev. Genet.* **19**, 770 (2018).
10. C. J. Smith *et al.*, bioRxiv 574020 (2019). <https://doi.org/10.1101/574020>.
11. M. J. Lajoie *et al.*, *Science* **342**, 357 (2013).
12. B. J. Karas *et al.*, *Chromosome Res.* **23**, 57 (2015).
13. L. A. Mitchell *et al.*, *Science* **355**, eaaf4831 (2017).
14. F. Ceroni, T. Ellis, *Nat. Rev. Mol. Cell Biol.* **19**, 481 (2018).

ACKNOWLEDGMENTS

We thank members of the GP-write technology and infrastructure working group for discussions. We also thank N. J. Kelley and A. Schwartz for administration and coordination of group discussions. B. J. K. is a cofounder, chief executive officer, and shareholder of Designer Microbes Inc. H. H. L. is a cofounder of Kern Systems. J. S. B. is a founder and director of Neochromosome, Inc. G. M. C.'s financial interests are listed at <http://arep.med.harvard.edu/gmc/tech.html>. J. W. E. is a cofounder and board member of Molecular Assemblies Inc. L. A. M. is a cofounder, president, and shareholder of Neochromosome, Inc. A. A. K. N. is a cofounder, chief executive officer, and shareholder of Asimov Inc. B. P. is a cofounder, chief technology officer, and shareholder of Twist Bioscience.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/310/suppl/DC1

10.1126/science.aay0339

ENVIRONMENTAL STUDIES

The work of words in the Anthropocene

Two books—one fictional, one dictionary—probe how we conceive of our changing world

By Deborah Dixon

The books reviewed here dwell on the work of words in an “Anthropocene,” a term itself critiqued as hubris-filled, which glosses the question of who is responsible for the forcing of physical processes and the role of the nonhuman in transforming a habitable planet. Both come from an arts and humanities tradition that worries at how we make sense of ourselves, the world, and our place within it—concerned that such contemplations are always partial, relational, and unfinished.

A growing environmental arts and humanities literature and practice has emphasized the emotional and psychological impacts of an Anthropocene that foregrounds a series of endings—from the extinction of species to the curtailing of life on Earth—as well as a series of challenges, from resilience to survival. What words do we use to describe the conditions we live in, the challenges we face, and the futures we want?

An Ecotopian Lexicon—“ecotopian” meaning out of place—offers a fascinating collection of non-English or newly invented words that impart something of the complexities of everyday life in an era of warming skies and oceans, mass degradation, precarity, and insecurity, each of which also helps map a possible future. The stated work of words here is to clarify, diagnose, and stimulate action.

Some of the terms are already circulating through academia and popular media. These include “solastalgia,” coined by Glenn Albrecht, which connotes a feeling of powerlessness and grief when faced with the sweeping change of a landscape that is still lived in, and “heyiya,” from Ursula Le Guin, which means to turn away from ecocidal practices. Some speak to an experience of place or its physical transformation (“godhuli” being a particularly evocative example, meaning refracted sunlight seen through the haze of dust kicked up by cows as they seek shelter). Some hint at earthly forces and cosmologies (such as “qi,” meaning a self-balancing order within the Universe that connects all and subsumes chaos). One greeting, written as “~*~,” connotes the feel of a light breath of air blown over the back of your hand; it is borrowed from dolphin societies to convey something of the electronic vibrations in which our bodies become enrolled.

This is a book that wants to stir passions, which in turn become a means of realizing desired futures. As such, it resonates with the Twitter work of so many climate scientists, which hinges on the relaying of dire facts in tones that urge action. The lexicon’s

“borrowing” of words, although mindful of a continued coloniality, nevertheless presents a welter of experiences, felt and articulated.

The words of *Extinction Events*, although similarly situated in “endings,” do a different kind of work. These words do not offer a map, nor do they seek to stir passions toward the future. Rather, they evoke, somewhat in the vein of Italo Calvino’s *The Complete Cosmicomics* (1), a planetary history that winds its way tightly into the lives of fictional women who face unsettling losses—from the disappearance of islands to the erosion of marital bonds.

What survives an extinction event? What emotional and psychological, as well as physical, traces remain? What does it feel like to anticipate cataclysm? And to experience the falling apart of maps that can no longer situate and orientate us? Breazeale’s stories press time and again on these questions, as the flesh of mother and daughter merges with the ash of volcanoes (“Ashcake”) and a memory of a kiss retains its Earthiness, even as it becomes a molecular cloud in the cosmos (“The Supernova of Irvin Edwards”).

At times, the subsuming of a human lively time by the “deep time” of the Earth is a little heavy-handed, but the best of this collection is profoundly affective. “Devil’s Tooth Museum,” for example, is a tour de force that ostensibly hinges on an effort to reopen an impact crater exhibition and museum center after the death of the protagonist’s sister, Syl. Syl’s presence haunts the site, prompting the girls’ grandfather to further the physical breakdown of the exhibits, even as the narrator strives to mend them. In the dusty, Lysol-scented atmosphere, the violent extinctions wrought by the impact merge with the final breakdown of a simulator and the loss of a sibling: “A cataclysm, the air vibrating. In the panels, across the glass cases, the swarm of colors perforating the dark and in the tremors Syl, her imprint a shattered cone, a remnant, a brokenness that inhabited the innermost hollows of the Earth.”

If *An Ecotopian Lexicon* makes futures with words, *Extinction Events* dwells on all of us as future fossils in the making. ■

An Ecotopian Lexicon

M. Schneider-Mayerson and B. R. Bellamy, Eds.
University of Minnesota Press,
2019. 344 pp.

Extinction Events

Liz Breazeale
University of Nebraska Press,
2019. 131 pp.



The word “godhuli” refers to the Sun, as seen through dust kicked up by cows.

The reviewer is at the School of Geographical and Earth Sciences, University of Glasgow, Glasgow G12 8QQ, UK. Email: deborah.dixon@glasgow.ac.uk

REFERENCES AND NOTES

1. I. Calvino, *The Complete Cosmicomics* (Houghton Mifflin Harcourt, 2014).

10.1126/science.aay3845

SOCIAL SCIENCE

The stories that make us spend (and save)

A Nobel laureate reveals how narratives shape the economy

By **Sonia Jaffe**

Where did boycotts originate? Who were the first Luddites? When were rolling suitcases invented? Stories such as these have traditionally garnered less attention from economists—who tend to focus on more easily quantified phenomena—than from anthropologists, sociologists, or psychologists. Robert Shiller wants to change that.

A Nobel laureate for his pioneering research on financial markets, Shiller argues in his new book, *Narrative Economics*, that the stories people tell can affect or even cause major economic events and therefore merit more attention from economists. By “narrative,” he refers to stories—and their underlying ideas or sentiments—that are transmitted through “word-of-mouth contagion,” including print, digital, and social media.

An analogy to epidemic dynamics runs throughout the book. Shiller references research from fields as diverse as marketing and neuroscience to describe how a narrative’s rate of “contagion” and “recovery” (how soon it is forgotten) are affected by human interest, celebrity, and a story’s visual imagery or ability to evoke intense emotion. Contagion and recovery rates, he argues, affect the prevalence and durability—the virality—of the associated narrative. Narratives “evolve” over time, potentially changing in their contagiousness and their effect on people’s behavior. “Mutations” arise when, for example, a slight change in wording or an association with a celebrity changes how contagious a narrative is; a small, random event can have compounding effects that lead to disproportionate influence.

Economic narratives—those that potentially affect people’s economic behavior—frequently include economic decisions that individuals who hear the story want to emulate. Shiller begins with the timely example of Bitcoin: If you hear that those who bought Bitcoin early got wealthy and its price is increasing, you will probably want to buy Bit-

coin. Such narratives are often self-fulfilling prophesies. Others may simply be false, with more accurate versions or explicit corrections attracting far less attention. True or false, Shiller argues that narratives can complement or reinforce each other and that a constellation of narratives sometimes affects events in ways that no single narrative would.

The bulk of Shiller’s discussion delves into nine narratives—or pairs of opposing narratives—that he identifies as having played important roles in the economy over the past 150 years, mainly of the United States and Europe. They range from concerns about technology eliminating jobs to tales of “evil businesses” raising prices to the competing



When we hear early Bitcoin investors got wealthy, we want to buy Bitcoin.

norms of conspicuous consumption and modest frugality. For each narrative, Shiller analyzes the rise and fall of salient phrases on Google Ngrams and in the ProQuest News and Newspapers database, but he focuses on the quotes and anecdotes that embodied the narratives at different points in history. The big events, such as the Great Depression and the Great Recession, will be familiar to most readers—the dozen other contractions and depressions since 1850 perhaps less so. For events big and small, the narratives Shiller presents based on contemporaneous primary sources are much more engaging than the employment, gross domestic product, and stock market numbers that economists generally use to describe them.

Narrative Economics: How Stories Go Viral and Drive Major Economic Events

Robert J. Shiller
Princeton University Press,
2019. 400 pp.



In addition to urging researchers to analyze the vast data on digital communication of narratives, Shiller advocates digitizing and assembling existing texts from the past and actively collecting data—through interviews and focus groups—about narratives of the present. He hopes that understanding the influence of narratives will allow leaders to “create and disseminate counternarratives that establish more rational and more public-spirited economic behavior.” He does not discuss the resemblance that such an intervention would have to government propaganda, although he does acknowledge the partisan associations of many narratives and the resulting challenges to objective research.

Shiller’s thorough discussion and many examples are certainly convincing as to the importance of narratives in individual economic decision-making and aggregate economic phenomena. However, most economists will likely want to focus on distinguishing narratives that cause economic events from those that are merely correlated with, or caused by, those events. Shiller acknowledges this challenge. Although the contagion process is “in many ways a random event,” there is no controlled randomization of narratives. Moreover, “big economic events usually can’t be described as caused by just a single constellation of narratives”; the interaction of competing or complementary narratives is an additional challenge for identifying causality.

Nevertheless, Shiller is optimistic about the potential for researchers to combine the growing quantities of available data with improvements in semantic analysis to provide the context necessary for distinguishing narratives’ causes from their effects. ■

The reviewer is at the Office of the Chief Economist, Microsoft, Redmond, WA 98052, USA.
Email: sonia.jaffe@microsoft.com

10.1126/science.aaz2690



LETTERS

Arid conditions limit the forest restoration potential of many regions of Australia.

Edited by Jennifer Sills

Forest restoration: Overlooked constraints

In their Report “The global tree restoration potential” (5 July, p. 76), J.-F. Bastin *et al.* use machine learning to derive the carbon storage potential of global tree restoration, which they identify as the most effective climate change mitigation option. However, the study likely overestimates the actual potential by identifying opportunities for increasing canopy cover in environments with obvious environmental or socioeconomic constraints.

In high-latitude regions of Russia, Scandinavia, and North America, permafrost and short growing seasons (1) impair tree growth. In large parts of Australia and other arid and hyperarid regions, salinity, sodicity, hardpans, and moisture limitations prevent tree establishment (2, 3). In African grasslands, infertile soils, grazing animals, water constraints, and wildfires maintain patchy shrub-grass environments (4). In areas with severely degraded soils and biodiversity loss in the Americas and in Asia (5, 6), prospects of restoring pre-degradation canopy cover are limited. In grazing lands and production forests, abandoning current uses implies staggering absolute opportunity costs. Finally, Bastin *et al.* excluded areas classified as urban, but the data set they used (7) fails to recognize some major

urban centers and many towns and villages in rural areas (7); more than 2.5 billion people live in areas that Bastin *et al.* considered eligible for restoration (8), including entire cities, such as Kinshasa, the capital of the Democratic Republic of Congo.

Bastin *et al.* introduced further overestimation by multiplying tree cover expansion potential by total ecosystem carbon. This operation lowers the baseline by assuming that carbon stock is proportionally related to canopy cover—i.e., that land with no trees contains no carbon. The use of biome-level carbon stock averages, without considering spatial variation, also adds considerable error, especially in alleged high-potential areas, where these averages (154.7 to 282.5 Mg ha⁻¹) are approximately 5 times greater than what has been reported in site-specific assessments (9, 10).

We appreciate the need for benchmark estimates of carbon storage and restoration potentials, but realistic predictions require tapping expert knowledge to ensure relevant constraints are considered, as well as more rigorous quality control, such as mapping how model validation errors are spatially distributed. Overly hopeful figures produced by models without necessary supervision may misguide the development of climate policy (11, 12).

Eike Luedeling¹, Jan Börner², Wulf Amelung³, Katja Schiffrers¹, Keith Shepherd⁴, Todd Rosenstock⁵

¹Department of Horticultural Sciences, Institute of Crop Science and Resource Conservation,

University of Bonn, 53121 Bonn, Germany.

²Department of Economics of Sustainable Land Use and Bioeconomy and Center for Development Research, Institute for Food and Resource Economics, University of Bonn, 53115 Bonn, Germany. ³Department of Soil Science and Soil Ecology, Institute of Crop Science and Resource Conservation, University of Bonn, 53115 Bonn, Germany. ⁴World Agroforestry Centre, Gigiri, Nairobi, Kenya. ⁵World Agroforestry Centre, Kinshasa, Democratic Republic of Congo.

*Corresponding author.

Email: luedeling@uni-bonn.de

REFERENCES AND NOTES

1. J. Obu *et al.*, *Earth-Sci. Rev.* **193**, 299 (2019).
2. FAO and ITPS, “Status of the World’s Soil Resources (SWSR)—Main Report: Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils” (Rome, Italy, 2015).
3. S. R. Morton *et al.*, *J. Arid Environ.* **75**, 313 (2011).
4. M. Sankaran *et al.*, *Nature* **438**, 846 (2005).
5. M. A. Stocking, *Science* **302**, 1356 (2003).
6. IPBES, “The IPBES assessment report on land degradation and restoration,” L. Montanarella, R. Scholes, A. Brainin, Eds. (Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany, 2018).
7. O. Arino, Global Land Cover Map for 2009, European Space Agency (ESA) & Université Catholique de Louvain (UCL), PANGAEA, 10.1594/PANGAEA.787668 (GlobCover 2009).
8. Center for International Earth Science Information Network, Columbia University, Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. (2018); <https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11/metadata>.
9. D. D. Shirima *et al.*, *African J. Ecol.* **49**, 332 (2011).
10. M. B. Siewert *et al.*, *J. Geophys. Res. Biogeosci.* **120**, 1973 (2015).
11. M. D. Mastrandrea *et al.*, “Guidance note for lead authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties” (IPCC, 2010); https://wg1.ipcc.ch/AR6/documents/AR5_Uncertainty_Guidance_Note.pdf.
12. C. M. Anderson *et al.*, *Science* **363**, 933 (2019).

10.1126/science.aay7988

Forest restoration: Expanding agriculture

In their Report “The global tree restoration potential” (5 July, p. 76), J.-F. Bastin *et al.* determine the available potential forest restoration area by excluding areas with existing trees, urban settlement, and cropland. However, they overestimate the potential area because they do not account for projected agricultural land expansion or current use of pasture land.

There is evidence from satellite imagery that most of global agricultural land expansion in the previous three decades happened and is still happening on tropical forest land, especially in Brazil and Southeast Asia (1–3). Given that this trend is likely to continue, especially in the highly productive areas in Central and South America, agricultural land expansion must be taken into account when assessing future tree restoration potentials (4–6). Food and Agriculture Organization projections expect an increase of cropland by 7% until 2030 (7), and evidence suggests an increase in global cropland area between 11 and 26% until 2050 (8), the latter corresponding to 4 million km². Based on one approach (4), not using this area for crop production would reduce global crop production by 11% and increase crop prices by 23%.

Furthermore, Bastin *et al.* assume that grassland would be available for tree restoration. They choose to ignore the data showing that currently about 30 million km² of grassland areas are used for extensive livestock production (9). Not utilizing areas for cropland expansion and pasture land requires a higher intensification of agriculture, which in turn is associated with higher agricultural emissions (10) and loss of biodiversity (4).

Bastin *et al.* do not consider current and future trade-offs with food security and neglect socioeconomic aspects of increasing consumption that arise through population growth, income growth, and preference changes toward more livestock products in fast-growing economies (11). Excluding estimated expansion areas and grazing land reduces the calculated sequestration potential by 19 and 57%, respectively, when applying the carbon densities of the book-keeping model BLUE (12).

Ruth Delzeit¹, Julia Pongratz^{2,3}, Julia M. Schneider², Franziska Schuenemann^{1*}, Wolfram Mauser², Florian Zabel^{1,2*}

¹Kiel Institute for the World Economy, 24105 Kiel, Germany. ²Ludwig-Maximilians-University, 80333 Munich, Germany. ³Max Planck Institute for Meteorology, 20146 Hamburg, Germany.

*Corresponding author. Email: franziska.schuenemann@ifw-kiel.de (F.S.); f.zabel@lmu.de (F.Z.)



Expanding agricultural land and livestock production could conflict with forest restoration goals.

REFERENCES AND NOTES

1. J. A. Foley *et al.*, *Nature* **478**, 337 (2011).
2. H. K. Gibbs *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16732 (2010).
3. Z. Zeng *et al.*, *Nat. Geosci.* **11**, 556 (2018).
4. F. Zabel *et al.*, *Nat. Commun.* **10**, 2844 (2019).
5. R. Delzeit, F. Zabel, C. Meyer, T. Václavík, *Region. Environ. Change* **17**, 1429 (2017).
6. A. Molotoks *et al.*, *Glob. Change Biol.* **24**, 5895 (2018).
7. N. Alexandratos, J. Bruinsma, “World agriculture towards 2030/2050: the 2012 revision,” *ESA Working Paper No. 12-03* (FAO, Rome, 2012).
8. C. Schmitz *et al.*, *Agric. Econ.* **45**, 69 (2014).
9. N. Ramankutty, A. T. Evan, C. Monfreda, J. A. Foley, *Glob. Biogeochem. Cycles* **22**, GB1003 (2008).
10. C. L. van Beek, B. G. Meerburg, R. L. M. Schils, J. Verhagen, P. J. Kuikman, *Environ. Sci. Pol.* **13**, 89 (2010).
11. F. Schuenemann, R. Delzeit, *Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V.* **64**, 185 (2019).
12. E. Hansis, S. J. Davis, J. Pongratz, *Glob. Biogeochem. Cycles* **29**, 1230 (2015).

10.1126/science.aaz0705

Forest restoration: Transformative trees

We welcome the attention given to forest and trees by the Report “The global tree restoration potential” (5 July, p. 76), in which J.-F. Bastin *et al.* study the potential of tree cover to reduce climate change. However, we are concerned by their neglect of the water cycle. They consider how water influences tree cover but disregard how tree cover influences water. Bastin *et al.* recognize that their extrapolations are not “future projections of potential forest extent” but instead represent potential tree cover “under existing environmental conditions.” However,

given the influence of forests on their environment, the concept of potential tree cover under current conditions is problematic. Trees influence several of the variables Bastin *et al.* used to model tree cover, including precipitation quantity, variability, and seasonality, as well as soil moisture and atmospheric water transport (1–4).

While much remains uncertain (2), we know enough to foresee that afforestation and reforestation have potential for both negative and positive hydrological impacts. Negative impacts can result if plantings deplete groundwater and thus exacerbate local water scarcity. Changes can manifest quickly and are a recognized problem with fast-growing monoculture plantations (5). Positive impacts can result when tree cover improves soil and groundwater recharge and storage, such as through suitable species and tree densities (6). Forest cover can also promote rainfall recycling and thus bolster and stabilize regional and downwind rainfall (1, 7, 8). In suitable circumstances, increased forest cover may even return wetter climates to currently drier regions, expanding the land available for trees (2). These outcomes have profound implications given that reliable access to water is central to achieving the UN Sustainable Development Goals. Accounting for the potentially transformative power of trees for both water and carbon offers crucial constraints as well as vast benefits.

Douglas Sheil,^{1,2} Aida Bargués-Tobella,^{3,4} Ulrik Ilstedt,⁴ Pierre L. Ibisch,⁵ Anastassia Makarieva,⁶ Clive McAlpine,⁷ Cindy E. Morris,⁸ Daniel



Murdiyaro,^{2,9} Antonio D. Nobre,¹⁰ Germán Poveda,¹¹ Dominick V. Spracklen,¹² Caroline A. Sullivan,¹³ Obbe A. Tuinenburg,¹⁴ Ruud J. van der Ent^{15,16}

¹Faculty of Environmental Sciences and Natural Resource Management (MINA), Norwegian University of Life Sciences (NMBU), 1432 Ås, Norway. ²Center for International Forestry Research (CIFOR), Bogor, Jawa Barat 16115, Indonesia. ³World Agroforestry Centre (ICRAF), 00100, Nairobi, Kenya. ⁴Department of Forest Ecology and Management, Swedish University of Agricultural Sciences (SLU), 901 83 Umeå, Sweden. ⁵Centre for Ecomics and Ecosystem Management, Faculty of Forest and Environment, Eberswalde University for Sustainable Development, D-16225, Eberswalde, Germany. ⁶Theoretical Physics Division, Petersburg Nuclear Physics Institute, St. Petersburg, Russia. ⁷School of Earth and Environmental Sciences, The University of Queensland, St. Lucia, QLD 4072, Australia. ⁸INRA, Plant Pathology Research Unit 407, PACA Research Center, 84143 Montfavet, France. ⁹Department of Geophysics and Meteorology, Bogor Agricultural University, Bogor, Indonesia. ¹⁰Centro de Ciencia do Sistema Terrestre INPE, São José dos Campos, São Paulo, Brazil. ¹¹Department of Geosciences and Environment, Universidad Nacional de Colombia, Medellín, Colombia. ¹²School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK. ¹³National Centre for Flood Research, Southern Cross University, Lismore, NSW 2480, Australia. ¹⁴Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, Netherlands. ¹⁵Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands. ¹⁶Department of Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, Netherlands. *Corresponding author. Email: douglas.sheil@nmbu.no

REFERENCES AND NOTES

1. D. Ellison *et al.*, *Glob. Environ. Change* **43**, 51 (2017).
2. D. Sheil, *For. Ecosyst.* **5**, 1 (2018).
3. J. S. Wright *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8481 (2017).

4. J. F. Salazar *et al.*, *Hydrol. Earth Syst. Sci.* **22**, 1735 (2018).
 5. R. B. Jackson *et al.*, *Science* **310**, 1944 (2005).
 6. U. Ilstedt *et al.*, *Sci. Rep.* **6**, 21930 (2016).
 7. L. Wang-Erlandsson *et al.*, *Hydrol. Earth Syst. Sci.* **22**, 4311 (2018).
 8. A. Staal *et al.*, *Nat. Clim. Change* **8**, 539 (2018).
- 10.1126/science.aay7309

Response

Luedeling and colleagues argue that we have overestimated the restoration capacity in several regions of the world. Our model predicts the expected optimal tree cover from a combination of 10 environmental variables that were selected through a variable selection procedure to avoid overfitting issues. As detailed in table S1 of our supplementary material, these 10 variables include mean annual temperature, temperature of the wettest quarter, annual precipitation, precipitation seasonality, precipitation of the driest quarter, elevation, hillshade, soil organic carbon, sand content, and depth to bedrock. These ecological variables cover average and seasonal variation in climate and variation in topographic and edaphic conditions. As such, we have done everything that is possible to represent all of the conditions raised by Luedeling and colleagues. Of course, cold and dry conditions are among the main limitations for tree growth, and that is why we have represented these environmental constraints in our model to ensure that we do not predict that trees can exist in regions that are too cold or dry.

As explained in the main text, our rigorous k-fold cross-validation (fig. S4A) revealed that our model could explain about 71% of the variation in tree cover without bias (fig. S3, B and C). This means that our model is unbiased at a global scale, but we do not explain 100% of the potential tree cover variation. It is consequently possible to find places where we overestimate or underestimate the potential tree cover—particularly in areas where uncertainties are high, as shown in fig S6.

Delzeit and colleagues claim that we overestimate the area available for tree restoration because the expansion of croplands in upcoming decades will reduce the land available for restoration, and because pasture lands are considered as potential land for restoration in our assessment. We agree that, if we continue to expand agricultural land area, there will be a reduction in the land available for restoration. As stated in our analysis, our model estimates the area that is currently available for restoration under present conditions. Of course, any changes in the area of land use will necessarily affect this

global total. We exemplified this in our attempts to show how future changes in climate might reduce the area available for restoration. We hope that our analysis can also serve as a stepping stone for future research to evaluate how changes in agricultural land use will affect the potential restoration area.

It is true that we included rangelands in the area available for restoration. Of course, much of this land is used for the grazing of animals and so may not be available for complete forest restoration. However, as mentioned in the Report, several studies suggest that it is possible to increase the current tree cover in these areas without limiting food production (1, 2), especially when forest cover is relatively low, as is the case for most of the pasture land in our model.

Because we removed all urban and agricultural land (i.e., we considered a potential increase of tree cover of 0% in cropland and urban areas), our numbers are likely to underestimate the total area that could currently be covered by trees. Indeed, both croplands and cities constitute great opportunities to increase the current tree cover and to play a major role in mitigating climate change (3–5). We maintain that our global estimate of the land available for restoration is a conservative one, and we encourage local land owners to use our forest restoration potential map in combination with more detailed local-scale estimates of land use when designing effective restoration strategies.

Sheil and colleagues point out that restoring ecosystems might have either positive or negative consequences regarding hydrology. We agree that these effects must be considered as a priority in upcoming research in restoration ecology.

Jean-Francois Bastin^{1*}, Yelena Finegold², Claude Garcia^{3,4}, Danilo Mollicone², Marcelo Rezende², Devin Routh¹, Constantin M. Zohner¹, Thomas W. Crowther¹

¹Crowther Lab, Department of Environmental Systems Science, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland. ²Food and Agriculture Organization of the United Nations, Rome, Italy. ³Department of Environmental Systems Science, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland. ⁴Centre de Coopération Internationale en la Recherche Agronomique pour le Développement (CIRAD), UR Forest and Societies, Montpellier, France.

*Corresponding author. Email: bastin.jf@gmail.com

REFERENCES AND NOTES

1. K.-H. Erb *et al.*, *Nature* **553**, 73 (2017).
2. K.-H. Erb *et al.*, *Nat. Commun.* **7**, 11382 (2016).
3. A. Albrecht, S. T. Kandji, *Agric. Ecosyst. Environ.* **99**, 15 (2003).
4. P. K. Ramachandran Nair, B. Mohan Kumar, V. D. Nair, J. Plant Nutr. Soil Sci. **172**, 10 (2009).
5. G. Manoli *et al.*, *Nature* **573**, 55 (2019).

10.1126/science.aaz2148

TECHNICAL COMMENT ABSTRACTS

Comment on “The global tree restoration potential”

Pierre Friedlingstein, Myles Allen, Josep G. Canadell, Glen P. Peters, Sonia I. Seneviratne Bastin *et al.* (Reports, 5 July 2019, p. 76) claim that global tree restoration is the most effective climate change solution to date, with a reported carbon storage potential of 205 gigatonnes of carbon. However, this estimate and its implications for climate mitigation are inconsistent with the dynamics of the global carbon cycle and its response to anthropogenic carbon dioxide emissions.
Full text: [dx.doi.org/10.1126/science.aay8060](https://doi.org/10.1126/science.aay8060)

Comment on “The global tree restoration potential”

Joseph W. Veldman, Julie C. Aleman, Swanni T. Alvarado, T. Michael Anderson, Sally Archibald, William J. Bond, Thomas W. Boutton, Nina Buchmann, Elise Buisson, Josep G. Canadell, Michele de Sá Dechoum, Milton H. Diaz-Toribio, Giselda Durigan, John J. Ewel, G. Wilson Fernandes, Alessandra Fidelis, Forrest Fleischman, Stephen P. Good, Daniel M. Griffith, Julia-Maria Hermann, William A. Hoffmann, Soizig Le Stradic, Caroline E. R. Lehmann, Gregory Mahy, Ashish N. Nerlekar, Jesse B. Nippert, Reed F. Noss, Colin P. Osborne, Gerhard E. Overbeck, Catherine L. Parr, Juli G. Pausas, R. Toby Pennington, Michael P. Perring, Francis E. Putz, Jayashree Ratnam, Mahesh Sankaran, Isabel B. Schmidt, Christine B. Schmitt,

Fernando A. O. Silveira, A. Carla Staver, Nicola Stevens, Christopher Still, Caroline A. E. Strömberg, Vicky M. Temperton, J. Morgan Varner, Nicholas P. Zaloumis Bastin *et al.*'s estimate (Reports, 5 July 2019, p. 76) that tree planting for climate change mitigation could sequester 205 gigatonnes of carbon is approximately five times too large. Their analysis inflated soil organic carbon gains, failed to safeguard against warming from trees at high latitudes and elevations, and considered afforestation of savannas, grasslands, and shrublands to be restoration.
Full text: [dx.doi.org/10.1126/science.aay7976](https://doi.org/10.1126/science.aay7976)

Comment on “The global tree restoration potential”

Simon L. Lewis, Edward T. A. Mitchard, Colin Prentice, Mark Maslin, Ben Poulter Bastin *et al.* (Reports, 5 July 2019, p. 76) state that the restoration potential of new forests globally is 205 gigatonnes of carbon, conclude that “global tree restoration [is] our most effective climate change solution to date,” and state that climate change will drive the loss of 450 million hectares of existing tropical forest by 2050. Here we show that these three statements are incorrect.
Full text: [dx.doi.org/10.1126/science.aaz0388](https://doi.org/10.1126/science.aaz0388)

Response to Comments on “The global tree restoration potential”

Jean-Francois Bastin, Yelena Finegold, Claude Garcia, Nick Gellie, Andrew Lowe, Danilo Mollicone, Marcelo Rezende, Devin Routh, Mactar Sacande, Ben Sparrow, Constantin M. Zohner, Thomas W. Crowther Our study quantified the global tree restoration potential and its associated carbon storage potential under existing climate conditions. We received multiple technical comments, both supporting and disputing our findings. We recognize that several issues raised in these comments are worthy of discussion. We therefore provide a detailed common answer where we show that our original estimations are accurate.
Full text: [dx.doi.org/10.1126/science.aay8108](https://doi.org/10.1126/science.aay8108)

Comment on “The global tree restoration potential”

Alan Grainger, Louis R. Iverson, Gregg H. Marland, Anantha Prasad Bastin *et al.* (Reports, 5 July 2019, p. 76) neglect considerable research into forest-based climate change mitigation during the 1980s and 1990s. This research supports some of their findings on the area of land technically suitable for expanding tree cover and can be used to extend their analysis to include the area of actually available land and operational feasibility.
Full text: [dx.doi.org/10.1126/science.aay8334](https://doi.org/10.1126/science.aay8334)

The Best Antibody Discovery Technology Is Now at Your Fingertips

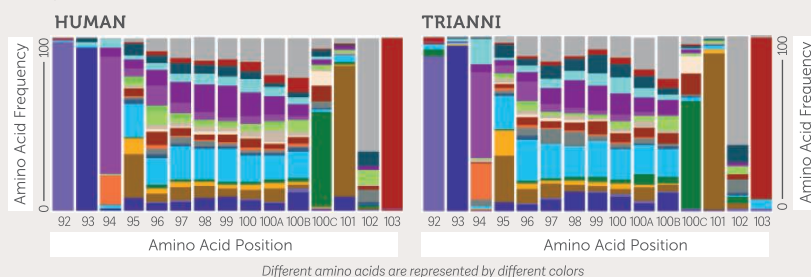
Trianni Mouse Antibodies are a Match for Humans

The Trianni Mouse™ platform is a transgenic antibody discovery platform offering the entire human variable gene diversity in a single organism.

The V(D)J gene segments in The Trianni Mouse are chimeric, but the variable domains of **antibodies made by the mouse are entirely human**. The result is human antibody leads generated from antibody genes optimized for function in the mouse.

To learn more about this innovative platform, visit Trianni.com.

Heavy Chain CDR3 Compositions of a Human Individual and a Trianni Mouse are Almost Identical



TRIANNI

Exceptional Human Antibody Discovery Technology

RESEARCH

IN SCIENCE JOURNALS

Edited by Michael Funk

EXOPLANETS

Peering inside extrasolar rocky bodies

The oxygen fugacity of a rock, f_{O_2} , is a measure of how oxidizing or reducing its surroundings were when the rock formed. Different minerals form at different f_{O_2} and have different physical properties, so the internal structure of an exoplanet depends on this value. Doyle *et al.* exploited the signature left behind when rocky bodies impact a white dwarf—the remnant of a dead star. By examining the rock-forming elements left on the

surface of each white dwarf, they determine f_{O_2} in the impacting body. Six systems all had similar f_{O_2} to bodies in the Solar System, consistent with the idea that rocky exoplanets often have internal properties similar to those of Earth and Mars. —KTS

Science, this issue p. 356

Artist's vision of a white dwarf stripping a planet of its rocky outer layers

CELL BIOLOGY

Origins of collective contraction

In contrast to plants and fungi, animals can deform their bodies by the collective activity of contractile cells. Collective contractility underlies processes such as gastrulation and muscle-based motility. Brunet *et al.* report that a close relative of animals, a choanoflagellate they name *Choanoeca flexa*, forms cup-shaped colonies that undergo collective contractility, leading to a rapid change in colony morphology (see the Perspective by Tomancak). *C. flexa* colonies are each composed of a monolayer of polarized cells. In response to sudden darkness, a light-sensing protein triggers coordinated, polarized contraction of *C. flexa* cells, which results in colony inversion. The

cellular mechanisms underlying this process are conserved between *C. flexa* and animals, indicating that their last common ancestor was also capable of polarized cell contraction. —BAP

Science, this issue p. 326;
see also p. 300

3D PRINTING

Large-scale, continuous 3D printing

Many three-dimensional (3D) printing methods build up structures layer by layer, which causes a lamination layer between each discrete step. Continuous printing can be done from a fluid bed if a so-called dead layer is used to buffer between the solidified structure and pool of resin. However, printing speeds are limited by the heat buildup from

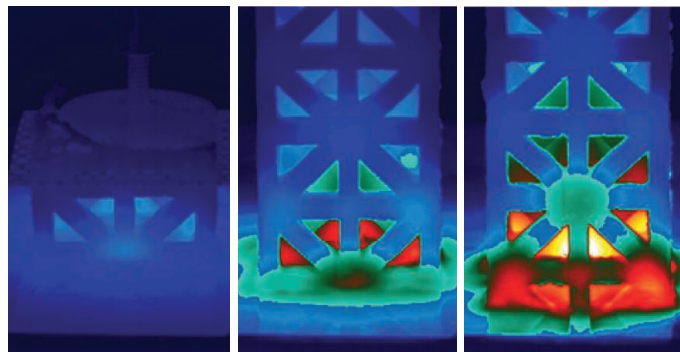
the exothermic polymerization process, thus limiting the ultimate size of the printed object. Walker *et al.* use a pumped, nonreactive fluorinated oil to act as the dead layer that removes heat during polymerization. This approach allows for both speedup and scale-up of the printing process. —MSL

Science, this issue p. 360

EARTHQUAKES

Many ruptures across many scales

The period of seismic quiescence in Southern California was rudely interrupted by the Ridgecrest earthquake sequence in July 2019. Ross *et al.* mapped the slip sequences during the magnitude 6.4 and



Thermal imaging of a 3D-printed lattice

7.1 earthquakes that shook the region. They found that ruptures of a few larger, but many smaller, faults occurred during both earthquakes. The Ridgecrest sequence calls for rethinking seismic hazard, as multifault ruptures are not usually considered when assessing seismic risk. —BG

Science, this issue p. 346

QUANTUM COMPUTING Generating large-scale cluster states

The development of a practical quantum computer requires universality, scalability, and fault tolerance. Although much progress is being made in circuit platforms in which arrays of qubits are addressed and manipulated individually, scale-up of such systems is experimentally challenging. Asavanant *et al.* and Larsen *et al.* explore an alternative route: measurement-based quantum computation, which is a platform based on the generation of large-scale cluster states. As these are optically prepared and easier to handle (one simply performs local measurements on each individual component of the cluster state), such a platform is readily scalable and fault tolerant. The topology of the cluster state ensures that the approach meets the requirements for quantum computation. —ISO

Science, this issue p. 373, p. 369

GENE EXPRESSION A statistical model to find disease genes

Genetic variation is high among individuals, which makes it difficult to identify any one specific pathogenetic variant in patients with idiopathic disease, especially those that are in non-coding regions of the genome. Examining tissue-specific and population-level RNA sequencing data, Mohammadi *et al.* developed a statistical test, analysis of expression variation (ANEVA), that can quantify how one individual's gene

expression fits in the context of the variation within the general population. By applying ANEVA to a dosage outlier test, the authors identified pathogenic gene transcripts in patients with Mendelian muscle dystrophy. —LMZ

Science, this issue p. 351

SLEEP Need for sleep

Sleep is crucial for healthy living and well-being, but individual sleep needs vary greatly. Xing *et al.* performed whole exome sequencing in a family of short sleepers and identified a point mutation in the neuropeptide S receptor 1 gene (*NPSR1*) responsible for the short sleep phenotype. The mutation increased receptor sensitivity to the endogenous ligand. Mice carrying the mutation showed increased mobility time, reduced sleep duration, and resistance to sleep deprivation-induced cognitive impairment. *NPSR1* may thus play a major role in sleep-related memory consolidation. —MM

Sci. Transl. Med. **11**, eaax2014 (2019).

FIBROSIS Sequencing in the matrix

Biological scaffolds that mimic tissue microenvironments can be used to model wound healing. Using single-cell RNA sequencing in two distinct biomaterial environments, Sommerfeld *et al.* found that defined populations of macrophages play a role in both fibrosis and regeneration. A subset of CD9⁺ interleukin-36 γ (IL-36 γ)-producing macrophages participate in IL-17-driven fibrosis. By evaluating wound healing in IL-17-deficient mice, they report IL-17 to be essential for the generation of CD9⁺ IL-36 γ -producing macrophages during fibrosis. Further studies are needed to understand the functional relationship between IL-17- and IL-36 γ -producing cells in fibrosis and in other settings. —AB

Sci. Immunol. **4**, eaax4783 (2019).

IN OTHER JOURNALS

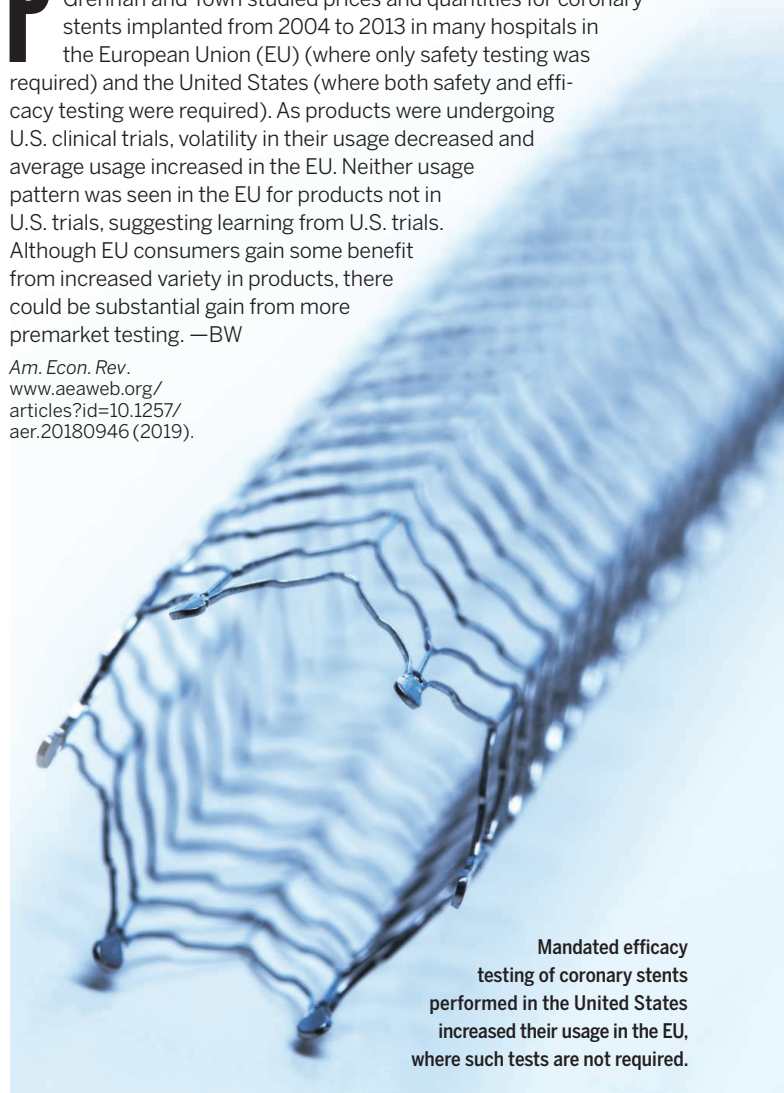
Edited by **Caroline Ash**
and **Jesse Smith**

MEDICAL REGULATION

The United States regulates, the European Union benefits

Premarket testing of medical products can be costly but can also reduce uncertainty and improve public welfare. Grennan and Town studied prices and quantities for coronary stents implanted from 2004 to 2013 in many hospitals in the European Union (EU) (where only safety testing was required) and the United States (where both safety and efficacy testing were required). As products were undergoing U.S. clinical trials, volatility in their usage decreased and average usage increased in the EU. Neither usage pattern was seen in the EU for products not in U.S. trials, suggesting learning from U.S. trials. Although EU consumers gain some benefit from increased variety in products, there could be substantial gain from more premarket testing. —BW

Am. Econ. Rev.
[www.aeaweb.org/
articles?id=10.1257/
aer.20180946](http://www.aeaweb.org/articles?id=10.1257/aer.20180946) (2019).



Mandated efficacy testing of coronary stents performed in the United States increased their usage in the EU, where such tests are not required.

ENERGY STORAGE Ultra-low-temperature aqueous batteries

The development of efficient electrochemical energy storage systems is of prime importance for future energy management, which will be increasingly based on renewable energy sources. Aqueous rechargeable alkali-ion batteries are particularly attractive for large-scale

implementations given their environmentally friendly components and low cost. However, their low-temperature performance requires improvements, which can be achieved by optimizing the electrolyte's ionic conductivity and freezing point. Using dimethyl sulfoxide as the electrolyte additive, Nian *et al.* develop a new practical electrochemical system with electrolyte that freezes below -130°C . As a



Studies on the baboon, *Papio anubis*, show how adverse life events affect subsequent generations.

LIFE HISTORY

Predicting offspring life span

Environmental and life choices affect health and reproduction. Several studies show that these effects extend to the next generation and sometimes beyond. Zippel *et al.* have examined a longitudinal dataset of a natural population of baboons in southern Kenya. Individuals in this group can live up to 30 years. The dataset was collected over 45 years and documented life span and adversity experienced during the prereproductive life of females. Adverse events ranged from death of the mother when offspring were young, to surviving drought, to having siblings of close age or being born into a large social group. Individuals affected by adversity often outlived their offspring. These older mothers seemed less able to successfully rear their offspring, who commonly experienced high mortality in infancy even when free of adversity. —BAP *eLife* **2019**, e47433 (2019).

result, the battery operates with sufficient ionic conductivity even at -50°C . The reasons for such an ultra-low freezing point are analyzed by means of spectral characterization and molecular dynamics simulation. —YS

Angew. Chem. Int. Ed. **10.1002/anie.201908913** (2019).

IMMUNOLOGY

Gallstones—a force of NETure?

Gallstones are a painful and sometimes fatal condition that is one of the leading causes of hospitalization in the United States. They are thought to form when bile is concentrated and acidified in the gallbladder, leading to the supersaturation and subsequent precipitation of cholesterol and calcium salts. Muñoz *et al.* report

that neutrophils release neutrophil extracellular traps (NETs) after interacting with calcium or cholesterol crystals. These NETs contain extracellular DNA and elastase, which help fuel crystallization and gallstone assembly. Mice with defects in pathways required for NET formation were found to have gallstones of reduced size and number. One future therapeutic target for this condition may therefore lie in preventing neutrophils from casting their NETs. —STS

Immunity **51**, 443 (2019).

PSYCHOLOGY

Community-based policing

Proponents of community-based policing argue that positive non-enforcement experiences with

police enhance citizens' trust and belief in the legitimacy of policing. Peyton *et al.* conducted a randomized field experiment to determine whether a single positive interaction could improve attitudes toward police. Police were trained in positive engagement strategies and assigned to talk with citizens for 10 minutes in a nonenforcement capacity. Surveys that were administered later found that citizens who had talked to officers held more positive attitudes toward the police 3 days and 21 days after those conversations compared with citizens who did not speak with an officer. These results have implications for improving relationships between citizens and police. —TSR

Proc. Natl. Acad. Sci. U.S.A. **116**, 19894 (2019).

MICROBIOLOGY

Globe-trotting conjugal hitchhiker

Plasmids transfer large gene clusters between distant bacterial relatives. They are notorious for spreading antibiotic resistance genes among pathogens. Petersen *et al.* wondered whether plasmid transfer among environmental bacteria is happening on a similar scale. Choosing *Roseobacter*, known for its genetic plasticity and importance in carbon and sulfur cycling in the oceans, the authors found a RepL-type plasmid called pLA6_12. This plasmid has 100% sequence identity in several genera from multiple locations. Consequently, its occurrence is indicative of recent global horizontal gene transfer events. pLA6_12 contains gene cassettes for detoxification of anthropogenic pollutants and, like antibiotic resistance plasmids, seems to be selected for by human activities. —CA

Proc. Natl. Acad. Sci. U.S.A. **116**, 20568 (2019).

BIOCHEMISTRY

Cooperating to control cadmium

Cadmium is toxic to living organisms because it displaces essential metal ions in macromolecules. Detoxification systems rely on sulfur-rich sites to coordinate the cadmium, but these sites also bind other heavy metals. Liu *et al.* show how the bacterium *Pseudomonas putida* eliminates cadmium. The protein CadR selectively responds to cadmium to bind to DNA and up-regulate transcription of itself and other cadmium detoxification proteins. The selectivity comes from cooperation between two types of binding sites. Cadmium first binds to a cysteine-rich site, and this stabilizes the second histidine-rich site in a conformation that supports cadmium binding. Binding at both sites facilitates DNA binding to activate transcription. —VV

Proc. Natl. Acad. Sci. U.S.A. **116**, 20398 (2019).

ALSO IN *SCIENCE* JOURNALS

Edited by Michael Funk

SIGNALING

Mediating systemic health

Sphingosine 1-phosphate (S1P) is an important circulating lipid mediator that is derived from the metabolism of cell membranes. Its diverse homeostatic roles, particularly in immunology and vascular biology, can go awry in numerous diseases, including multiple sclerosis, cardiovascular diseases, and fibrosis. The centrality of S1P signaling has led to the development of several drugs, including two approved for treatment of multiple sclerosis. In a Review, Cartier and Hla discuss the current understanding of how one mediator can carry out so many signaling roles in different tissues, how these become dysregulated in disease, and efforts in drug development to target S1P signaling. —GKA

Science, this issue p. 323

HUMAN EVOLUTION

Adaptive archaic hominin genes

As they migrated out of Africa and into Europe and Asia, anatomically modern humans interbred with archaic hominins, such as Neanderthals and Denisovans. The result of this genetic introgression on the recipient populations has been of considerable interest, especially in cases of selection for specific archaic genetic variants. Hsieh *et al.* characterized adaptive structural variants and copy number variants that are likely targets of positive selection in Melanesians. Focusing on population-specific regions of the genome that carry duplicated genes and show an excess of amino acid replacements provides evidence for one of the mechanisms by which genetic novelty can arise and result in differentiation between human genomes. —LMZ

Science, this issue p. 324

NEURODEVELOPMENT

Retinal neurons play musical chairs

As the retina develops in fruit flies, subtypes of certain color photoreceptors, the R7 neurons, are stochastically specified even though the downstream neurons they need to connect to, the Dm8 neurons, are not randomly specified. Courgeon and Desplan find that Dm8 neurons are actually specified in subtypes and are produced in excess. Those Dm8 neurons that connect to an R7 input survive; those that fail to find a match die by apoptosis. The matching is facilitated by a pair of cell adhesion molecules. The stochastic results of R7 differentiation are thus propagated downstream even though those downstream elements are not stochastically determined. —PJH

Science, this issue p. 325

ORGANIC CHEMISTRY

Charging through the looking glass

Asymmetric catalysis is a commonly applied technique to prepare just one of two mirror-image products in a chemical reaction. But what if you already have the compound you want, stuck in a mixture of left- and right-handed enantiomers? Shin *et al.* now show that light-induced electron transfer can trigger a favorable succession of proton and hydrogen-atom transfer steps, both of which are susceptible to biasing by catalysts, to preferentially convert a mixture of cyclic urea enantiomers into just one (see the Perspective by Wendlandt). —JSY

Science, this issue p. 364;
see also p. 304

BIODIVERSITY

Spatial structure of species change

Biodiversity is undergoing rapid change driven by climate

change and other human influences. Blowes *et al.* analyze the global patterns in temporal change in biodiversity using a large quantity of time-series data from different regions (see the Perspective by Eriksson and Hillebrand). Their findings reveal clear spatial patterns in richness and composition change, where marine taxa exhibit the highest rates of change. The marine tropics, in particular, emerge as hotspots of species richness losses. Given that human activities are affecting biodiversity in magnitudes and directions that differ across the planet, these findings will provide a much needed biogeographic understanding of biodiversity change that can help inform conservation prioritization. —AMS

Science, this issue p. 339;
see also p. 308

NEUROSCIENCE

Special moments at cortical quiet states

Delta waves are moments of widespread cortical silence that alternate with active states during slow-wave sleep. However, upon closer examination, single neuronal action potentials can be detected during delta waves. Todorova and Zugaro sought to determine whether this neuronal noise could instead be an important signal (see the Perspective by Ikegaya and Matsumoto). They found that persisting action potential firing during delta waves is an overlooked but widespread phenomenon, which could potentially involve all neurons and all delta waves. A critical role of the delta wave may be to insulate specific cortical computations taking place in response to hippocampal replay and involved in memory consolidation. —PRS

Science, this issue p. 377;
see also p. 306

TURBULENCE

Down and down the energy cascade

Injecting energy into a turbulent system at large length scales results in the energy cascading down and eventually dissipating at a characteristic small length scale. In conventional fluids, this small scale is set by fluid viscosity. Navon *et al.* studied the turbulence energy cascade in a quantum gas, a Bose-Einstein condensate of rubidium-87 atoms held in a uniform trap. Dissipation occurred by atoms escaping from the trap at a scale that could be tuned by varying the height of the trapping potential. Thanks to the flexibility of their setup, the researchers were able to study both the steady state, in which energy is injected at the same rate it is dissipated, and the transient regime preceding the steady state. —JS

Science, this issue p. 382

WATER OXIDATION

Inspecting S states in photosynthesis

Oxygenic photosynthesis uses a Mn_4CaO_5 cluster in the oxygen-evolving complex to extract electrons from water and produce dioxygen. Visualizing each of the chemical states in this process, S_0 to S_4 , and assigning chemical identities and mechanisms on the basis of structures has been a challenge addressed recently by work at x-ray free-electron lasers. Suga *et al.* used serial crystallography at cryogenic temperatures to trap and determine the structures of several stable states during photosystem II water oxidation (see the Perspective by Britt and Marchiori). Changes around the water cluster already happen in the S_2 state and set the stage for water insertion that occurs during transition to the S_3 state. A short 1.9-angstrom distance between the two oxygen atoms in the S_3 state is consistent with

theoretical studies supporting an oxyl/oxo mechanism for oxygen-oxygen coupling. —MAF

Science, this issue p. 334;
see also p. 305

IMMUNOLOGY

Mobilizing T cells

Accumulating evidence suggests that a nonclassical antigen presentation pathway, mediated by human leukocyte antigen-E (HLA-E), has an important role in regulating innate and adaptive immune responses to infections and cancer. In a Perspective, Ottenhoff and Joosten discuss the dichotomous role of HLA-E in suppressing immune responses (through an immune checkpoint), as well as in activating unconventional T cells. This nonclassical pathway has the potential to be manipulated to prevent infectious disease through vaccination and to treat cancer through immunotherapy. —GKA

Science, this issue p. 302

NEURODEVELOPMENT

How GABA makes the switch

A high concentration of chloride ions in neurons interferes with signaling mediated by the neurotransmitter γ -aminobutyric acid (GABA). The transporter KCC2 extrudes chloride ions, and Watanabe *et al.* and Pisella *et al.* examined mice expressing KCC2 with mutations that mimicked constitutive phosphorylation at two regulatory sites (see the Focus by Zamponi). Dephosphorylation of these sites in KCC2 enables GABA to switch from being an excitatory to being an inhibitory neurotransmitter during neurodevelopment, thus allowing the maturation of the neurocircuitry underlying cognition and respiration. —LKF

Sci. Signal. **12**, eaaw9315, eaay0300, eaay8960 (2019).

NEUROSCIENCE

MORE model to fight against addiction

The United States is in the midst of an opioid crisis affecting more than 15 million Americans. Garland *et al.* provide neurophysiological evidence for a therapeutic approach to help individuals with opioid use disorder (OUD). They conducted a series of randomized experiments using an approach called Mindfulness-Oriented Recovery Enhancement, or MORE. Hedonic dysregulation in brain reward circuitry has been viewed as a core mechanism of addictive behavior; the MORE model aims to restructure reward responsiveness using mindfulness techniques. MORE may remediate hedonic dysregulation by simultaneously increasing responsiveness to natural rewards and decreasing reactivity to drug-related rewards. The use of MORE and other cognitive training interventions may ultimately serve to turn the tide on addiction in the United States. —KSL

Sci. Adv. 10.1126/sciadv.aax1569 (2019).

REVIEW SUMMARY

SIGNALING

Sphingosine 1-phosphate: Lipid signaling in pathology and therapy

Andreane Cartier and Timothy Hla*

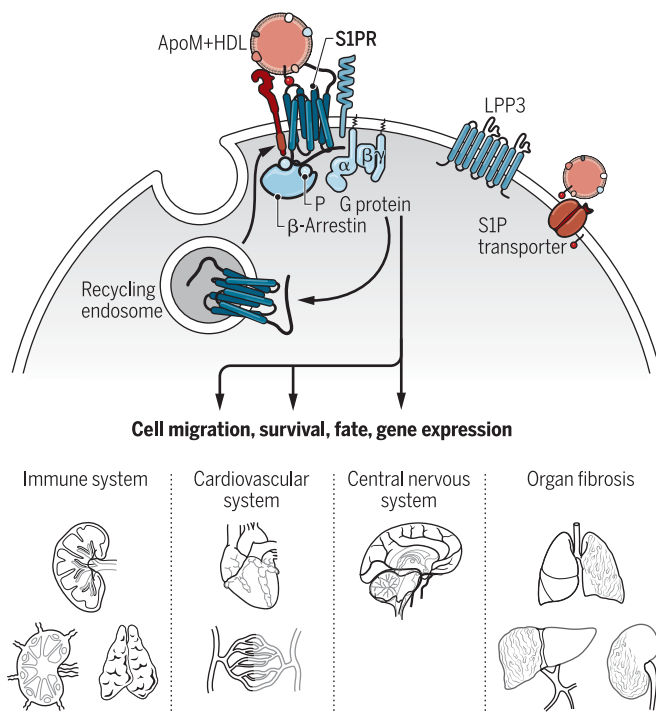
BACKGROUND: Sphingosine 1-phosphate (S1P), a product of membrane sphingolipid metabolism, is secreted and acts through G protein-coupled S1P receptors (S1PRs) in vertebrates. S1PR isoforms mediate complex cellular actions either alone or in combination in most organ systems. This stable lysolipid circulates as a complex with protein chaperones that not only enables aqueous solubility but also helps facilitate specific modes of receptor signaling. However, differential concentration gradients of S1P are normally present in various compartments and are perturbed under disease conditions. The abundance of circulatory S1P and the high expression of S1PRs in exposed cells—that is, vascular and hematopoietic cells—poses a key question of how this signaling axis is regulated. This question is of clinical relevance because the first S1PR-targeted drug, fingolimod, has been approved for the treatment of multiple sclerosis since 2010. Recent findings from basic research as well as insights gleaned from clinical and translational studies have enriched our understanding of how this simple lysolipid evolved as a complex regula-

tor of multiple physiological systems and, when dysregulated, contributes to numerous diseases.

ADVANCES: Extracellular spatial gradients of S1P, demonstrated by using S1P reporters, are tightly regulated and control fundamental processes such as hematopoietic cell trafficking, immune cell fate, and vascular integrity. The gradients are formed through location-specific function of metabolic enzymes, S1P transporters, and chaperones. Such physiological S1P gradients are altered in diseases, thus contributing to conditions such as inflammation, autoimmunity, and vascular dysfunction. S1P complexed to chaperone proteins—for example, high-density lipoprotein-bound apolipoprotein M—mediate distinct modes of receptor activation, resulting in biased receptor signaling and specific biological outcomes. S1PRs are also regulated tightly through endocytic mechanisms and receptor modulators that enhance or inhibit signal strength and duration. Various signaling mechanisms of this simple lysolipid mediator has helped reveal its multiple actions in the immune system, which include

S1PR signaling regulates multiorgan pathophysiological processes.

Extracellular S1P gradients created by transporters, chaperones (ApoM⁺HDL), and metabolic enzymes (LPP3) interact with S1PRs on the cell surface. Receptor activity, transmitted by means of G proteins, is regulated by multiple mechanisms, including β -arrestin coupling, endocytosis, and receptor modulators. The resultant cellular changes influence multiple organ systems in physiology and disease.



adaptive immune cell localization in various compartments (egress versus retention), fate switching, survival, and activation that influences both cell-mediated and humoral immunity. In the cardiovascular system, high expression of multiple S1PR isoforms in various cell types regulate development, homeostasis, and physiology. Current S1PR-targeted drugs that aim to tame autoimmunity exhibit con-

siderable cardiovascular-adverse events. In the central nervous system (CNS), widespread application of S1PR-targeted drugs in autoimmune neuroinflammatory diseases has stimulated research that revealed the broad but poorly understood effects of S1P signaling in neurodevelopment, the neurovascular unit, neurons, and glia. Furthermore, in addition to the involvement of pathological S1P signaling in acute ischemic conditions of various organs, chronic dysregulated S1P signaling has been implicated in fibrotic diseases of lung, heart, liver, and kidney.

OUTLOOK: Considerable challenges remain to fully harness the new knowledge in S1P pathobiology to translational utility in clinical medicine. Approaches that mimic S1P chaperones, S1P neutralizing agents, modulation of transporters, biased agonists and antagonists of S1PR isotypes, and sphingolipid metabolic enzyme modulators provide viable pathways to therapy. Focusing on the immune system, such approaches may widen the autoimmunity therapeutic landscape and provide new directions in cancer and chronic inflammatory diseases. For cardiovascular diseases, ischemic conditions as well as chronic heart failure are likely candidates for future translational efforts. Although further work is needed, S1P-targeted approaches may also be useful in regenerative therapies for the aging and diseased myocardium. The CNS-targeted efforts may cross into neurodegenerative diseases, given the success with S1PR-targeted drugs in reducing brain atrophy in multiple sclerosis. Other potential applications include approaches in pain management and neurodevelopmental disorders. Such strategies, although challenging, are greatly helped by findings from basic research on S1P pathobiology as well as pharmacological and clinical insights derived from the application of S1P-targeted therapeutics. ■

Vascular Biology Program, Boston Children's Hospital and Department of Surgery, Harvard Medical School, Boston, MA 02115, USA.

*Corresponding author. Email: timothy.hla@childrens.harvard.edu

Cite this article as A. Cartier and T. Hla, *Science* 366, eaar5551 (2019). DOI: 10.1126/science.aar5551

REVIEW

SIGNALING

Sphingosine 1-phosphate: Lipid signaling in pathology and therapy

Andreane Cartier and Timothy Hla*

Sphingosine 1-phosphate (S1P), a metabolic product of cell membrane sphingolipids, is bound to extracellular chaperones, is enriched in circulatory fluids, and binds to G protein-coupled S1P receptors (S1PRs) to regulate embryonic development, postnatal organ function, and disease. S1PRs regulate essential processes such as adaptive immune cell trafficking, vascular development, and homeostasis. Moreover, S1PR signaling is a driver of multiple diseases. The past decade has witnessed an exponential growth in this field, in part because of multidisciplinary research focused on this lipid mediator and the application of S1PR-targeted drugs in clinical medicine. This has revealed fundamental principles of lysophospholipid mediator signaling that not only clarify the complex and wide ranging actions of S1P but also guide the development of therapeutics and translational directions in immunological, cardiovascular, neurological, inflammatory, and fibrotic diseases.

Lipids are best known for their roles in energy storage and formation of cellular membranes. The diversity of membrane lipid structures ensures structural flexibility and integrity of cellular membranes needed for adaptation to varying environments. As vertebrate evolution ensued, metabolites of membrane lipids were repurposed as extracellular ligands for G protein-coupled receptors (GPCRs), which in turn activate numerous intracellular signaling pathways, thus coupling membrane phospholipid breakdown with intercellular communication (1). This coupling was first illustrated by platelet activating factor, a bioactive lipid synthesized from membrane phospholipids of hematopoietic and vascular cells that activates GPCRs in an autocrine and paracrine manner to evoke allergic reactions and anaphylaxis (2). Similarly, sphingosine 1-phosphate (S1P)—which is produced and secreted by red blood cells (RBCs), endothelial cells, and platelets into the extracellular environment—acts on specific GPCRs expressed by most if not all cell types (3).

S1P was discovered in the 1960s as a terminal product of sphingolipid metabolism (4). Studies in cultured cells suggested its bioactive nature, even though the prevailing notion at the time considered it more akin to classical second messengers such as diacylglycerol and Ca^{2+} (5, 6). We now know that vertebrates possess five S1P receptors (S1PR1 to S1PR5) that respond to extracellular S1P to regulate embryonic development, physiological homeostasis, and pathogenic processes in multiple organ systems (3). These high-affinity S1PRs couple

to key intracellular signaling pathways controlled by heterotrimeric G proteins, Rho family small guanosine triphosphatases (GTPases), and the protein kinase AKT, resulting in changes in cytoskeleton, cell adhesion, and survival (7). Although the extracellular actions of S1P are well established, distinct intracellular functions of S1P have also been suggested. For example, S1P was purported to bind to and regulate the activities of cytosolic signal transduction mediator tumor necrosis factor- α (TNF- α) receptor-associated factor 2 (TRAF-2) (8), the chromatin-modifying enzyme histone deacetylase-1 (HDAC-1) (9), the mitochondrial regulator prohibitin-2 (10), atypical protein kinase-C (11), and the catalytic subunit of telomere reverse transcriptase (12). Whether intracellular S1P signaling is physiologically relevant is not clear because genetic studies do not support the essential or functional roles of intracellular S1P targets (13–16).

The distinct physicochemical properties of S1P endow its multifunctional nature in developmental, physiological, and pathological contexts (17). For example, resistance to oxidative modification makes S1P suitable as a modulator of inflammatory processes, which are characterized by the influx of cells with high oxidative capacity. Poor aqueous solubility of S1P necessitates its association with protein molecules for transport in the blood and lymph circulation and access to S1PRs (18). Such interactions also facilitate the formation of S1P gradients, which are spatial concentration differences across biological compartments, thus allowing selective receptor activation at various locations (18–21). These properties enable S1P to be involved in multiple physiological processes. The essential nature of S1P functions has been illustrated in humans with congenital conditions such as Sjögren-Larsson

syndrome, nephrosis, adrenal insufficiency, and hearing impairment and in mouse genetic studies in which major embryologic and postnatal defects in various organs were observed when S1PRs, transporters, and S1P metabolic enzymes were mutated (3, 22, 23).

S1P-regulated processes contribute to several cardiovascular, autoimmune, inflammatory, neurological, oncologic, and fibrotic diseases (3). Moreover, the successful U.S. Food and Drug Administration (FDA) approval of an S1PR-targeted drug, fingolimod, to treat multiple sclerosis (MS) in 2010 (24) has led to considerable growth of research and drug development efforts. However, fingolimod binds and activates four S1PRs (S1PR1, -3, -4, and -5) while also functionally antagonizing S1PR1. Recently, a more specific S1PR1- and S1PR5-targeted drug, siponimod, was approved for the treatment of MS. In this Review, we discuss how conceptual advances in understanding S1P signaling have reached a critical threshold to allow rational design of new therapeutic strategies.

S1P spatial gradients

Extracellular gradients of growth factors, cytokines, and developmental morphogens are critical for achieving fidelity and specificity of biological responses (25). Lipid mediators, small molecules with limited aqueous solubility but affinity for hydrophobic molecules, adopted distinct signaling mechanisms to overcome these physicochemical challenges. For example, tissue- and cell-specific secretion, chaperone association, and extracellular metabolism of S1P result in the formation of S1P spatial gradients that are important for its numerous biological activities.

The enzymes that generate S1P—namely, ceramidase and sphingosine kinase (SPHK)—do double duty in intracellular sphingolipid metabolism as well as extracellular signaling (26). In many tissues, S1P is rapidly degraded by the endoplasmic reticulum-resident S1P lyase, resulting in very low intracellular S1P concentrations. By contrast, S1P can also be coupled to export processes by means of specific transporters. These flip the S1P molecule from the inner to the extracellular leaflet of the plasma membrane, allowing extracellular action of S1P (3). Once exported, S1P is capable of activating S1PRs expressed in the same cell in an autocrine manner. However, extraction by chaperone proteins, which bind to S1P and enable aqueous solubility, lead to diffusible S1P in the extracellular environment, thus creating spatial gradients that activate S1PRs in a paracrine and/or endocrine manner (Fig. 1).

Tissue-specific transporters that belong to the “major facilitator” superfamily export S1P to establish extracellular gradients. For example, spinster homolog 2 (SPNS2), first identified in a zebrafish screen for cardia bifida (a

Vascular Biology Program, Boston Children's Hospital and Department of Surgery, Harvard Medical School, Boston, MA 02115, USA.

*Corresponding author. Email: timothy.hla@childrens.harvard.edu

developmental condition that leads to the formation of two beating hearts), functions in the extraembryonic yolk syncytial layer to export S1P, which binds to S1PR2 on the cells of the endoderm, which in turn allows directional cardiac progenitor cell migration to the midline and proper cardiac development (27). This is the first example of S1P extracellular gradients created by a lipid transporter that is required for a developmental event (28, 29). In mice, SPNS2 exports S1P from vascular and lymphatic endothelial cells (LECs) to activate S1PRs on lymphocytes, thus allowing their egress from lymph nodes into the circulatory system (30, 31). High circulating S1P concentrations together with low amounts of S1P in interstitial fluid in secondary lymphoid organs, such as lymph nodes, are critical for lymphocyte egress into lymphatic circulation (19).

Mouse models that use reporters to sense extracellular S1P revealed S1P gradients at higher resolution. In the spleen, extracellular S1P gradients formed by the phosphatase enzyme lipid phosphate phosphohydrolase 3 (LPP3) are essential for marginal zone B cell trafficking (32), whereas LEC-expressed SPNS2 in lymph nodes regulated medullary cord natural killer (NK) cell location and function (33). The transporter major facilitator superfamily domain-containing protein 2B (MFSD2B), which exports S1P from RBCs into plasma, is also important for their normal function and turnover (34). Loss of function of S1P transporters leads not only to developmental defects but also to early onset hearing loss (35) and retinal defects (36). These observations suggest that spatial S1P signaling, which is critical in embryogenesis and postnatal physiological processes, leads to pathological conditions when dysregulated. Cellular mechanisms by which S1P transporters are controlled are poorly understood. Strategies to restore normal

transporter function could be useful in the treatment of various diseases in which extracellular S1P gradients are perturbed.

Chaperones, biased signaling, and S1PR regulators

S1P chaperone proteins, which stably bind and transport S1P in circulatory and interstitial fluids, facilitate S1PR activation on recipient cells. The prototypical S1P chaperone is apolipoprotein M (ApoM), a component of high-density lipoprotein (HDL), a circulating lipoprotein associated with vascular health (20). In addition, serum albumin binds to S1P at lower affinity (37). HDL-bound ApoM binds S1P stably in blood plasma and activates S1PRs on endothelial cells to suppress inflammation, promote barrier function, allow liver regeneration after hepatectomy, and protect from bacterial endotoxin-induced inflammatory responses (38–41). In hematopoietic progenitor cells, ApoM-bound S1P suppresses lymphopoiesis (42). HDL-bound S1P activates S1PR1 as a biased agonist that selectively induces certain biological responses. Other HDL-bound factors—such as apolipoprotein A1, cellular HDL receptor scavenger receptor-B1, and the ability of HDL to modulate membrane raft-based organization of receptor-effector complexes—may all be involved in biased signaling (43). Additionally, the relative stability of HDL-bound S1P compared with albumin-bound S1P may also play a role in differential kinetics of receptor activation and signaling (Fig. 2) (38). By contrast, albumin-bound S1P is short-lived in plasma (44) and activates S1PRs in a manner distinct from that of HDL-S1P.

Once activated by small-molecule agonists or albumin-bound S1P, S1PR expression on the cell surface is down-regulated, which makes vascular and hematopoietic cells exposed to circulatory S1P refractory to further stimula-

tion. Regulation of S1PR1 endocytosis is precisely controlled by GPCR kinase 2, dynamin, and moesin (45–48). Loss of such S1PR1 regulatory events leads to changes in biological outcomes such as altered immune cell trafficking (46) and vascular endothelial barrier function (49). Sustained endocytosis of lymphocyte S1PR1 explains the therapeutic efficacy of S1PR1 functional antagonists in the treatment of autoimmune diseases.

The recent development of S1PR reporter mice has revealed plasma S1P-dependent β -arrestin activation in vascular endothelium upon systemic endotoxemia (50, 51) and in the aortic endothelium at areas of disturbed flow (38). This suggests that S1PR1 is down-regulated in injured and inflamed endothelium. The ability to monitor S1PR function in vivo in real time and at single-cell resolution is anticipated to lead to new insights into how S1P signaling is regulated physiologically and in disease.

In addition to the natural ligand S1P, other molecules as well as coreceptors modulate S1PR activity. For example, conjugated bile acids (CBAs) activate S1PR2, which may be important in hepatobiliary and intestinal systems during cholestatic injury (52). In endothelial cells, activated protein C receptor and the hyaluronic acid receptor CD44 transactivate S1PR1 to increase barrier function and thereby prevent plasma leakage, thus limiting inflammatory responses (53, 54). In lymphocytes, activation-induced CD69 binds to S1PR1 and down-regulates its expression, an event that is critical for regulation of lymphocyte egress and tissue residence (55). In the central nervous system (CNS), the binding of the extracellular domain of Nogo-A, a neuronal repulsive molecule that regulates neural network formation, to S1PR2 induces signals by means of the GPCR Gα13-RhoA signaling pathway to repress neurite outgrowth, thus achieving synaptic plasticity in the hippocampus and motor cortex (56). An unbiased, genome-wide CRISPR/Cas9 screen for S1PR1 modulators identified lysophosphatidic acid receptor-1 as a negative regulator. This axis is important in lymphatic endothelial cells of the lymph node sinuses (57). In $\gamma\delta$ T cells, S1PR2 inhibition of S1PR1 function restrains them in the skin, thus promoting immunity in the local environment by gaining tissue residency (58).

These examples illustrate that S1PRs evolved to interact with many other signaling pathways involved in embryonic development, inflammation, host defense, and homeostasis. Mechanistic details of how S1P transfer takes place from the chaperones to S1PRs, the action of various receptor modulatory molecules, and the functions of regulated endocytosis and plasma membrane retention of S1PRs need to be further elucidated. Interestingly, the crystal structure of agonist-bound S1PR1 suggested

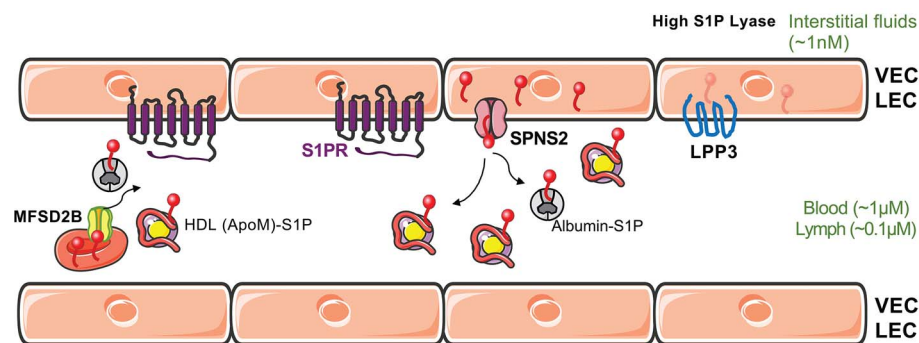
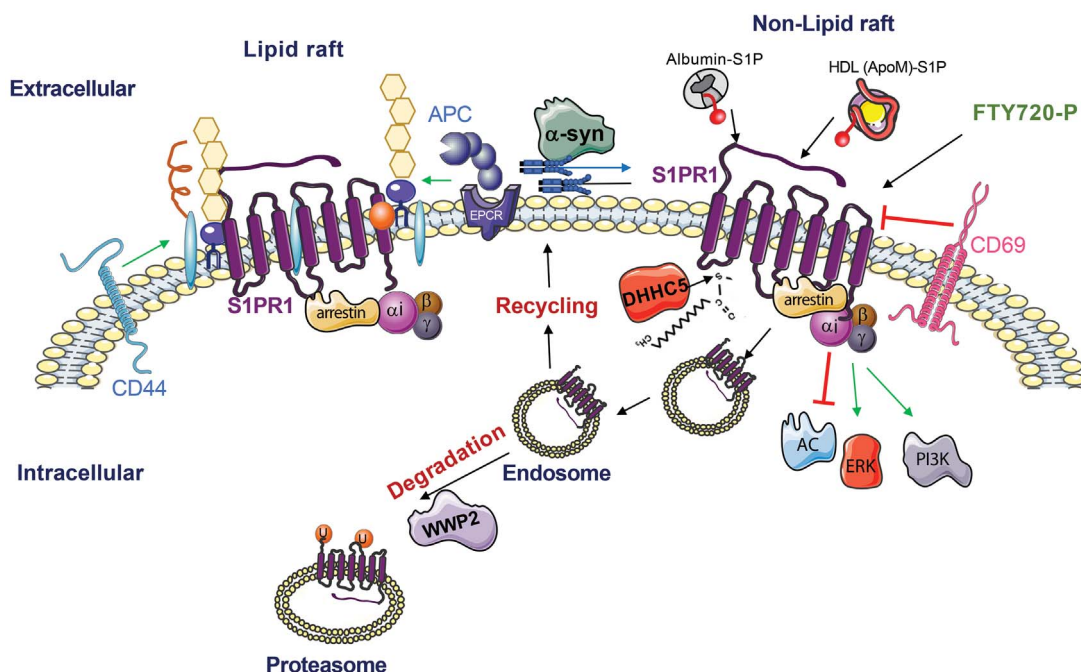


Fig. 1. Establishing and maintaining S1P gradients in vessels. S1P is transported by chaperones ApoM-HDL and albumin in the circulation and presented to S1PRs. S1P export by vascular endothelial cells (VECs) and lymphatic endothelial cells (LECs) (by SPNS2) and by RBCs (by MFSD2B) as well as degradation of S1P by the phosphatase LPP3 modify the extracellular S1P gradient. Low interstitial tissue S1P is achieved by the S1P lyase. Tissue concentrations of S1P are indicated in parentheses. S1PR, sphingosine 1-phosphate receptor; SPNS2, sphingolipid transporter 2; MFSD2B, major facilitator superfamily domain containing 2B; LPP3, lipid phosphate phosphatase 3.

Fig. 2. S1PR1 signaling regulation.

Schematic representation of regulators of S1PR1 cell surface expression, signaling, and turnover. Albumin-S1P, HDL-S1P, and FTY720-P activate S1PR1. Internalization of S1PR1 eventually leads to proteasomal degradation by the ubiquitin ligase WW domain containing E3 ubiquitin protein ligase 2 (WWP2). CD44 and activated protein C (APC)/endothelial protein C receptor (EPCR) transactivate S1PR1, while CD69 binds to and down-regulates S1PR1 at the cell surface. Endocytosis of S1PR1 is achieved through GPCR kinase 2 (GRK2)-dependent phosphorylation and endocytic regulators (dynamin, moesin, β -arrestin, and AP2/Clastrin). The intracellular signaling of S1PR1 is transmitted by means of the heterotrimeric Gai family proteins, the inhibition of adenylyl cyclase, and activation of phosphatidylinositol 3-kinase (PI3K)-AKT and ERK pathways. Palmitoylation by DHHC5 affects the coupling of S1PR1 to Gai, whereas α -synuclein modulates the coupling of S1PR1 with Gai by removing S1PR1 from lipid rafts. Regulators of other S1PRs are not well characterized.



that ligand (S1P) entry takes place by means of the transmembrane domains of S1PR1 (59). Therefore, it is likely that HDL docking on the plasma membrane enables efficient S1P transfer to S1PRs by way of a transmembrane route. New structural analysis methods, such as cryo-electron microscopy, may reveal dynamic receptor conformations that are anticipated to lead to insights for the development of highly specific receptor modulatory agents. This may also lead to general understanding of how lipid mediators function in regulating cellular and organ function.

Cardiovascular development and pathophysiology

The vasculature, one of the first organ systems to develop in vertebrate embryogenesis, is regulated by circulatory S1P signaling (Fig. 3) (60, 61). Initial phases of angiogenesis encompass hypoxia-induction of vascular endothelial growth factor (VEGF) that induces endothelial sprouts, in which S1P signaling is minimal. When vascular sprouts fuse to form primary vascular networks, circulatory S1P accesses endothelial cells. These fusion events coincide with the induction of S1PR1 expression in the endothelium, and consequently, the formation of adherens junctions and cell-extracellular matrix adhesions that enable the vascular barrier to form (62, 63). Activation of endothelial nitric oxide synthase (eNOS) dilates vessels and allows proper blood flow (61). Multiple redundant sources of circulating S1P from RBCs, en-

dothelial cells, and platelets and the existence of sensing mechanisms to monitor and maintain circulating S1P levels during vascular development reinforces the essential nature of spatial S1P gradients (64, 65).

Postnatally, S1P signaling regulates normal vascular homeostasis. For example, S1P-S1PR1 signaling maintains endothelial cell barrier functions to limit plasma and leukocyte extravasation (66, 67). Deficiency in this homeostatic signaling system, such as reduced HDL-S1P, leads to compromised barrier function and vascular tone, which is needed to control blood flow and systemic blood pressure (39, 68). In addition, S1P signaling in the endothelium protects the glycocalyx, a protective, glycoprotein-rich layer that lines the apical surface of blood vessels (69). Recent identification of the endoplasmic reticulum-resident protein, Nogo-B, as a negative regulator of de novo sphingolipid synthesis and autocrine S1P signaling in the endothelium suggests the presence of local, as well as systemic, factors in the regulation of vascular tone and blood pressure (70, 71). The control of local S1P signaling may be vascular bed-specific—that is, in arteries versus veins—and possibly even organ-specific to enable precise control of the vascular system in response to metabolic and physiological cues of a specific organ system (72).

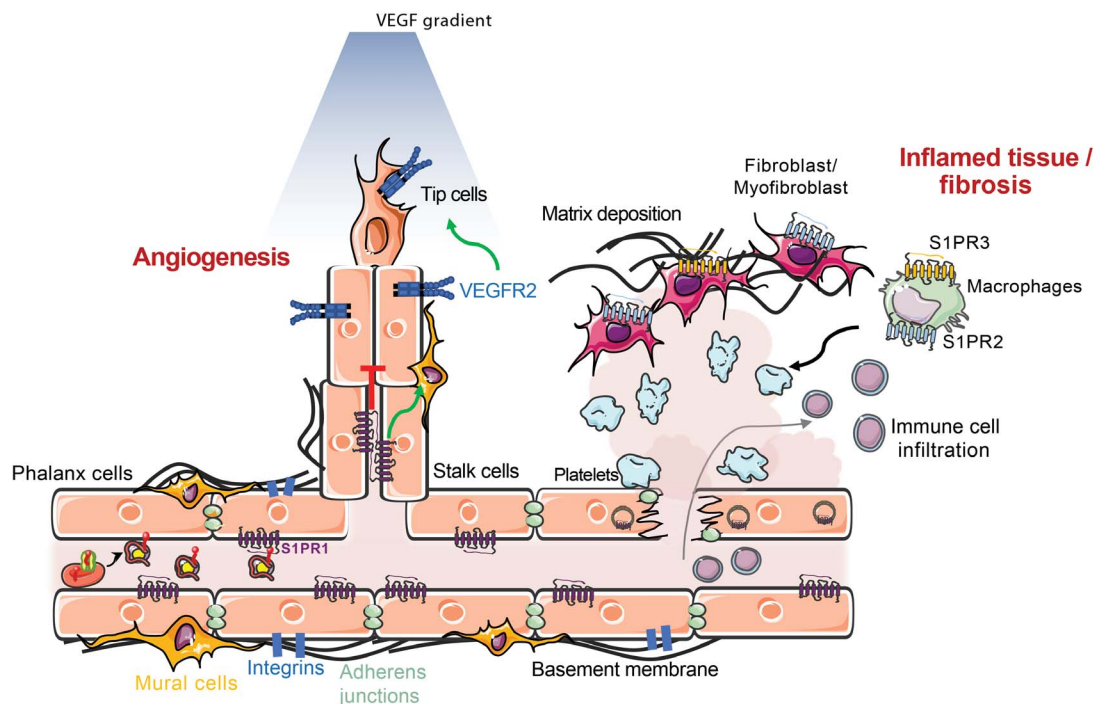
Diabetes and hypertension as well as premature birth frequently lead to vascular diseases of the eye (retina), such as retinopathy, which is characterized by abnormal blood flow,

excessive vascular leak, and blood vessel overgrowth. Neutralizing S1P antibodies reduced retinal neovascularization, and *S1pr2*-deficient mice showed attenuated angiogenesis in a mouse model of retinopathy of prematurity, a condition that occurs in premature neonates exposed to high levels of oxygen (73, 74). In diabetes as well as retinopathy of prematurity, the motility of Müller glial cells—which monitor retinal homeostasis, support neurons, and contribute to retinal structure and function—is regulated by S1P activation of S1PR3, suggesting a role for pathologic S1P signaling in retinal astrogliosis (scar formation) (75). S1P signaling may be a useful therapeutic target for proliferative retina diseases, such as diabetic retinopathy and age-related macular degeneration (wet form).

Cardiac development is also regulated by S1PR signaling. In zebrafish, the SPHK2-SPNS2-S1PR2-hippo signaling pathway is essential for endoderm development and myocardial progenitor cell migration (14, 28, 76). In mice, S1PR1 function in the endocardium, cardiomyocytes, and endothelial cells is essential for embryonic heart development (77). In human induced pluripotent stem cells, S1P induces cardiomyocyte differentiation (78), suggesting a critical role of S1P in development of the myocardium (heart muscle). Although S1P pathway alterations have not been causally linked with congenital heart defects in humans, these observations suggest that S1P is involved in postnatal cardiac remodeling events, including

Fig. 3. S1P function in vascular development, homeostasis, and pathology.

Vascular network development, maturation, and stabilization are regulated by endothelial S1PR1. Activation of S1PR1 inhibits sprouting angiogenesis and vascular endothelial growth factor receptor 2 (VEGFR2) and promotes vascular stability through adherens junctions and cell–extracellular matrix adhesion (integrins). Loss of cell-surface S1PR1 and destabilization of the endothelium causes the leakage of plasma and immune cells into the tissue, leading to inflammation. Vascular damage also leads to thrombosis (platelet aggregation) and fibrin deposition in the interstitial space, which activates fibroblasts and macrophages. Loss of S1PR1 signaling on endothelial cells and activated S1PR2 and S1PR3 signaling on macrophages and fibroblasts ultimately lead to tissue fibrosis, with extracellular matrix deposition, and organ dysfunction.



heart failure, a complex disease in which cardiac tissue is remodeled and the pumping efficiency of the heart diminishes.

Postnatally, cardiac S1PRs regulate several pathophysiological processes such as cardiac rhythm, vascular perfusion, and myocardial protection. S1P induces bradycardia (slow heart rate) by activating the S1PR1-regulated G protein-coupled inwardly-rectifying potassium (GIRK) channels, which hyperpolarize myocardial cells and reduce the rate of action potential firing (79). HDL-induced S1P signaling by means of endothelial S1PRs allows efficient coronary circulation through eNOS-dependent vasodilatation and protection from myocardial ischemia-reperfusion injury, which is an important clinical problem after heart attacks (80). Importantly, the S1P pathway is cardioprotective. During ischemic preconditioning, a phenomenon in which small and transient ischemia protects from massive ischemic insults, SPHK is induced in the myocardium, resulting in enhanced synthesis of S1P, which protects the heart by promoting circulation and reducing the heart rate (81). Further, in a porcine model of ischemia-reperfusion injury, activation of S1PRs by fingolimod reduced infarct size and improved cardiac function (82). Moreover, in a mouse model of acute myocardial infarction, treatment with an engineered S1P chaperone (ApoM-Fc) suppressed cardiac injury and enhanced the recovery of cardiac function (67).

An endogenous regulator of S1P metabolism, Nogo-B was shown to modulate plasma and

inflammatory cell extravasation from blood vessels in pressure-overloaded hearts and pathological cardiac hypertrophy through endothelial S1PR1 autocrine signaling (83). The complex developmental and pathophysiological roles of S1P in the cardiovascular system provides therapeutic opportunities. For example, endothelial cell-targeted S1PR1 agonism may be useful in limiting tissue damage during ischemia-reperfusion injury. In addition, down-regulating S1PRs in autoimmune diseases leads to adverse cardiovascular effects, such as bradycardia and vascular leakage, which were predicted by pre-clinical and mechanistic studies (84). Therefore, S1PR-targeted drugs that are more selective for either the heart, vascular, or immune S1PRs may provide improved safety profiles in the control of cardiovascular and autoimmune disorders, respectively.

Complex actions in immunity

In addition to its established roles in immune cell trafficking, S1P signaling regulates diverse immunological processes, including cell fate switching, cell survival, innate immunity, and antitumor immune responses (Fig. 4).

Recent studies have illustrated the role of S1P gradients and S1PR signaling in the migration and retention of adaptive immune cells in resident tissues, thus influencing the specificity and magnitude of immune responses. Endothelial cell SPNS2, by means of the export of S1P, mediates the egress of immature B cells from the bone marrow and naïve T cells from the thymus (30, 31). In the spleen, mature B cell

localization and shuttling at the specific immunological region of this organ, called the marginal zone (MZ), is critical for the adaptive immune response against blood-borne antigens that get trapped in the spleen (85). B cell S1PR1 ensures MZ localization, whereas S1PR3 inhibits MZ shuttling, both of which are important for optimal immune responses (86).

B cell confinement and growth regulation in the germinal centers of lymphoid follicles, which are sites for B cell maturation and clonal selection, is controlled by S1PR2 signaling through the Gα13 GPCR signaling pathway (87). In the absence of this spatial restriction, B cell mislocalization is observed and may contribute to the dissemination of lymphoma (a B cell tumor). A recent study showed that in glioblastoma brain tumor patients, naïve T cell S1PR1 is down-regulated, leading to sequestration of such cells in bone marrow, resulting in lymphopenia and decreased antitumor immune responses (88). Prevention of this tumor-induced S1PR1 loss of function in T cells could enhance antitumor immunotherapy treatments. These illustrative examples demonstrate how spatial S1P gradients and cell-specific S1PR isotypes shape adaptive immune responses.

CD8⁺ resident memory T cells (CD8⁺T_{RM} cells)—long-lived, noncirculating cells that are critical for immunological memory and host defense—show reduced expression of the transcription factor Kruppel-like factor 2 (KLF2) and its downstream target gene *S1pr1* in mice, thus effectively minimizing egress and achieving

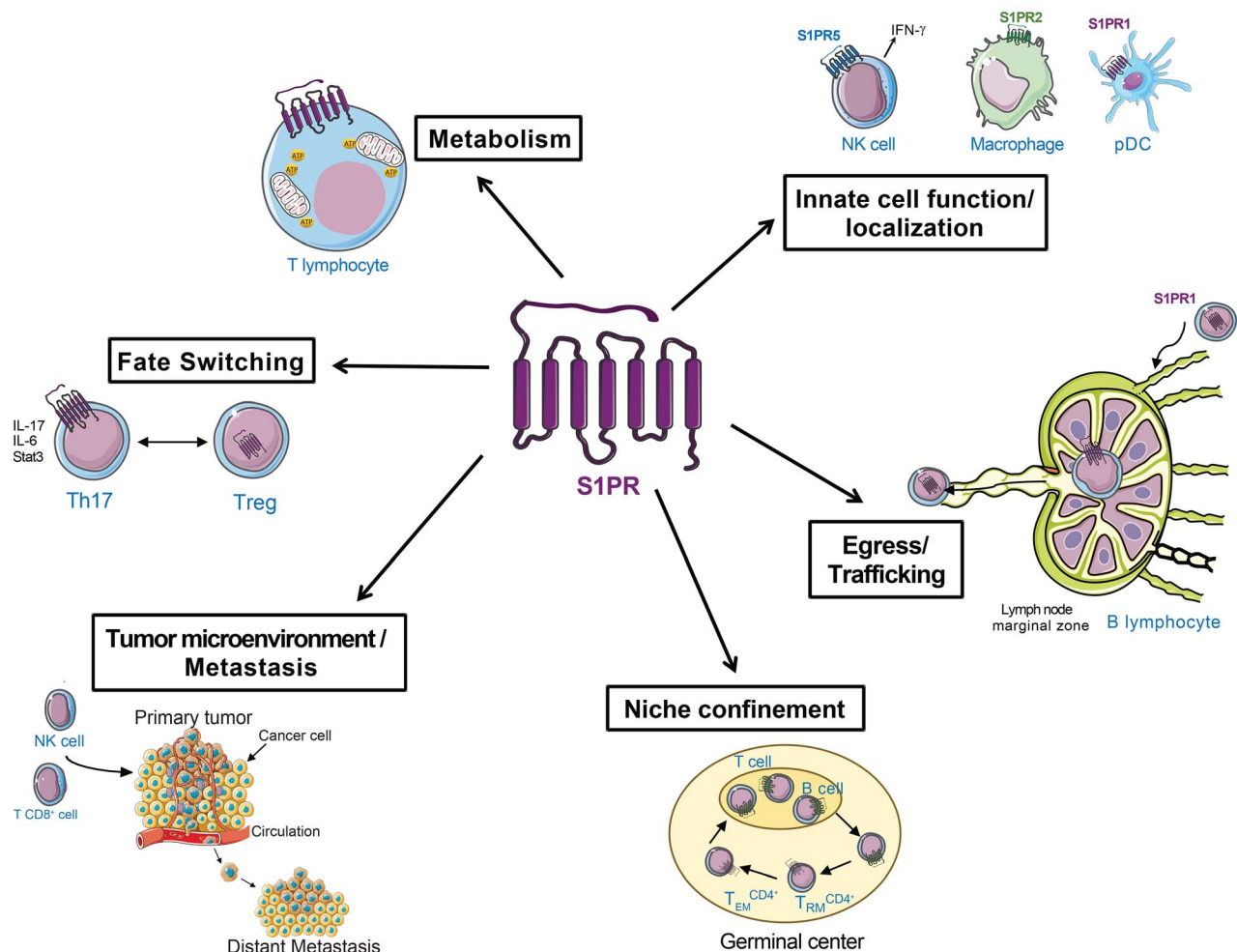


Fig. 4. S1P function in immunity. Immune cell function, egress, and survival are highly dependent on the regulation of S1P and S1PR signaling. Innate immune cell localization to inflammatory sites and egress of B and T cells from bone marrow and thymus, toward and out of secondary lymphoid organ marginal zones, rely on S1PR cell surface expression and S1P gradient. The metabolism of T cell ATP as well as the regulation of fate switching between T_H17 and regulatory T cells are also S1PR1-dependent. The confinement of activated CD4⁺ T cells in germinal centers within secondary lymphoid organs and spleen is regulated by S1PR2, whereas SPNS2 and S1PR1 are shown to be implicated in the recruitment of effector T cells to the tumor microenvironment and suppression of metastasis.

tissue retention (89). In addition, the cell-surface lectin CD69 down-regulates S1PR1 in T cells, which also blocks tissue egress and reinforces T_{RM} cell differentiation. CD69 expression is also critical for tissue retention of not only T_{RM} cells but also early effector T cells, thus enabling rapid immunological memory responses and the magnitude of the adaptive immune response (90). The migration of human effector memory T cells (which are circulating T_{RM} cells) and recently activated T cells is inhibited by S1P because of the resulting down-regulation of S1PR1 and enhanced expression of S1PR2 on these cells (91). Thus, S1PR-dependent signaling in various memory T cells shape long-term immunological memory, a key defense mechanism against repeated infections.

In addition, innate immune cell localization and function, which are important for host defense and inflammatory processes, are regulated by S1PR signaling. The positioning of NK

cells in the medulla of lymph nodes, which is important for rapid production of the inflammatory cytokine interferon- γ (IFN- γ) after infection, is regulated by S1PR5 (33). S1PR5 expression is induced during NK cell maturation and regulates the egress of NK cells from the bone marrow and secondary lymphoid organs (92). In plasmacytoid dendritic cells, S1PR1 blunts the amplification of inflammatory IFN- α -dependent downstream immune responses (93). Intestinal type-2 innate lymphoid cells undergo interorgan migration, a property that is essential for their protective role in parasitic infections, in a S1P-dependent manner (94). S1PR2 promotes the retention of macrophages at sites of inflammation by inhibiting chemokine-induced cellular motility (95). The generality of S1PR-dependent immune cell trafficking and retention mechanisms in influencing pathological processes was recently illustrated by the discovery that the S1P trans-

porter SPNS2 is essential for the recruitment of antitumor CD8⁺ T cells and NK cells to the tumor microenvironment, resulting in the inhibition of metastasis (96).

Moreover, S1PRs also modulate immune cell energy metabolism, which modulates cell fate during immune responses. Naïve T cells constantly survey lymphoid organs for foreign antigens, which necessitates high metabolic flux to generate intracellular adenosine triphosphate (ATP). In LECs, SPNS2 exports S1P, which activates lymphocyte S1PR1 and helps maintain the mitochondrial content of naïve T cells, thus ensuring high ATP generation by mitochondria (97). S1P is also necessary for the persistent activation of CD4⁺ T cells in inflamed tissues (98).

S1P signaling in specific contexts appears to regulate T cell subtype fate switching, which determines divergent immune responses. S1PR1 phosphorylation and subsequent internalization

in T cells regulates T helper 17 (T_H17) cell polarization, which induces pro-inflammatory immune reactions. Specifically, S1PR1 on the plasma membrane of immature T cells promotes differentiation into T_H17 cells by means of interleukin-6 (IL-6)-dependent signal transducer and activator of transcription 3 (STAT3) signaling and enhances autoimmune neuroinflammation (99). *S1pr1*-deficient mice and MS patients treated with fingolimod (which down-regulates S1PR1), exhibit increased levels of regulatory T cells (T_{reg} cells), which suppress inflammatory responses (100). Thus, S1PR1 regulates proper T_H17 and T_{reg} cell distribution across peripheral organs and thereby influences the magnitude of adaptive immune responses. Permanent deletion of *S1pr1* in T_{reg} cells leads to spontaneous autoimmunity, whereas acute deletion of *S1pr1* induces a higher inflammatory response in the experimental autoimmune encephalomyelitis (EAE) model, an animal model for MS in which adaptive immune cells enter the CNS to induce myelin destruction (101). A key unexplored question is how altered signaling of S1PR1 translates to T cell fate switching. It is possible that altered T cell localization brought about by S1PR1-dependent trafficking and retention mechanisms might change the magnitude of cytokine responses that induce fate switching. Alternatively, S1PR signaling may influence T cell-intrinsic mechanisms that induce fate switching—forkhead box P3 (FOXP3) for T_{reg} cells and retinoid-related orphan receptor- γ (ROR γ) and STAT3 for T_H17 cells. However, modulation of the transcription factor STAT3 expression may be involved in S1PR1-dependent pathological immune reactions (102, 103). Although many mechanistic steps have been uncovered, much needs to be learned about receptor specificity, redundant signaling, and involvement in normal and pathological immune responses. This could allow rational design of immunological and anti-inflammatory therapeutics that target S1P signaling.

The central nervous system

The neurovascular unit of the CNS forms a selective barrier called the blood-brain barrier (BBB), which is composed of endothelial cells, pericytes, astrocytes, and microglia and protects the brain parenchyma from circulatory elements. Dysregulated S1P signaling disrupts the BBB, which is an early event that contributes to many CNS diseases, including MS, Alzheimer's disease (AD), and ischemic stroke (Fig. 5) (104). Brain endothelial cell S1PR1 regulates tight junction protein functions, thus leading to the establishment of the BBB. Studies in mice revealed that loss of this mechanism resulted in a size-selective [molecular weight < 3 kD] breach of the BBB (105), presumably because of defective tight junction function

and increased leakage by way of a paracellular pathway (that is, between cells). Because transient pharmacological inhibition of brain endothelial cell S1PR1 leads to increased CNS penetration of small molecules, this approach may potentially be useful in enhancing drug delivery in neurological diseases (105). In addition, P glycoprotein, which exports small molecules out of the CNS into the circulation, is inhibited by S1PR inhibitors, which allows enhanced penetration of small-molecule drugs into the CNS (106). Such preclinical studies have sparked interest in inhibiting S1PR1 to enhance CNS-targeted therapeutics, a major bottleneck in modern medicine.

In acute ischemic stroke, SPHK2-dependent S1P production in the brain tissue is neuroprotective, presumably because of extracellular signaling by S1PRs (107). By contrast, S1PR2, which is induced in vascular endothelial cells after ischemia, was associated with increased infarct size, edema, cerebrovascular permeability, hemorrhage, and neurovascular injury (108). Fingolimod was effective in rat models of intracerebral hemorrhage and stroke through immunological and nonimmunological mechanisms (109). These preclinical studies have prompted the initiation of clinical trials to use fingolimod in the treatment of stroke (110). These studies also highlight the S1P axis in protecting the vasculature during ischemic stress and hence the underlying neuronal tissue after stroke.

The expression of S1PR2 was higher in vascular endothelium and astrocytes within neuroinflammatory disease-susceptible CNS regions of an inbred strain (SJL) of female mice and in the white matter of female MS patients than in those of males. In mouse EAE models, S1PR2 function is responsible for the breakdown of adherens junctions, BBB leakage, and chemokine-dependent inflammatory responses, thus contributing to disease severity (111). Given that autoimmune neuroinflammatory diseases, such as MS, occur more commonly in females, the identification of S1PR2 as a sex-dimorphic regulator of the BBB permeability raises a number of interesting questions. For example, what are the mechanisms by which blood vessels in the brain of females express higher levels of S1PR2? Would targeting this receptor provide a therapeutic opportunity for the treatment of MS, especially in treatment-resistant female MS patients?

In MS, autoreactive immune cells enter the CNS and induce inflammation, leading to demyelination. In addition to S1PR1 activation in autoreactive lymphocytes, which is an important target of S1P therapeutics, S1PR1 and S1PR3 expressed by astrocytes may influence neuroinflammatory processes (112, 113). Recent studies suggest that SPHK1 and S1PR3 are associated with inflammatory phenotypes of glia (114, 115). More defined mouse genetic studies are needed

to precisely elucidate S1PR-mediated inflammatory processes in both immune cells and resident microglial cells of the CNS in various forms of MS. This is an important issue to resolve for the development of more effective therapeutics.

Recent studies also suggest the involvement of S1PR signaling in the pathogenesis of AD, Parkinson's disease (PD), and Huntington's disease (HD). Associations between decreased S1P levels with AD (116, 117), PD (118), and HD (119) were observed, especially in early stages of these diseases. In addition, defective S1PR signaling was seen in cell culture (120) and preclinical models of neurodegenerative disease (121). Given the high unmet medical need for these conditions, the prominent signaling of S1PRs in the CNS suggests that better understanding of the exact roles of S1P signaling in AD, PD, and HD could lead to new therapeutic approaches.

The pathogenesis of fibrosis

Chronic tissue damage coupled with reduced resolution of inflammatory and wound healing responses lead to fibrosis (scarring), thus impairing normal organ function. Abnormal S1P signaling plays an important role in the fibrosis of multiple organ systems (Fig. 3). When the microvasculature is injured or dysfunctional, tissue parenchyma becomes flooded with plasma S1P, which leads to the activation of S1PRs in interstitial cells. In addition, platelet activation and thrombosis also cause increased local secretion of S1P, which leads to abnormal S1PR activation. Attenuated vascular S1PR1 and increased parenchymal and immune cell S1PR2 and S1PR3 leads to activation of Rho GTPase and Hippo-Yes-associated protein (YAP) signaling pathways that drive fibrosis (76). Thus, pathological S1P signaling in parenchymal and infiltrating immune cells promotes fibrotic diseases of the kidney, liver, and lung.

Acute kidney injury caused by perioperative renal ischemia and reperfusion leads to damage and dysfunction of the tissue, resulting in up-regulation of S1PR2 expression in renal proximal tubule cells, which promotes kidney fibrosis. By contrast, SPHK1 and S1PR1 signaling attenuated renal injury, presumably by salvaging the vasculature (122). Treatment of mice with fingolimod decreased the infiltration of immune cells and fibrosis in the kidney after unilateral urethral obstruction (UUO), a model of kidney injury, likely because of transient activation of vascular S1PR1 and inhibition of immune cell trafficking to the kidney (123). In tubulointerstitial renal fibrosis induced in mice, SPHK2 is up-regulated. *Sphk2*-deficient mice show less severe fibrosis in the UUO model, and this correlates with suppression of transforming growth factor- β (TGF- β) signaling in the kidney parenchymal cells and increased macrophage polarization

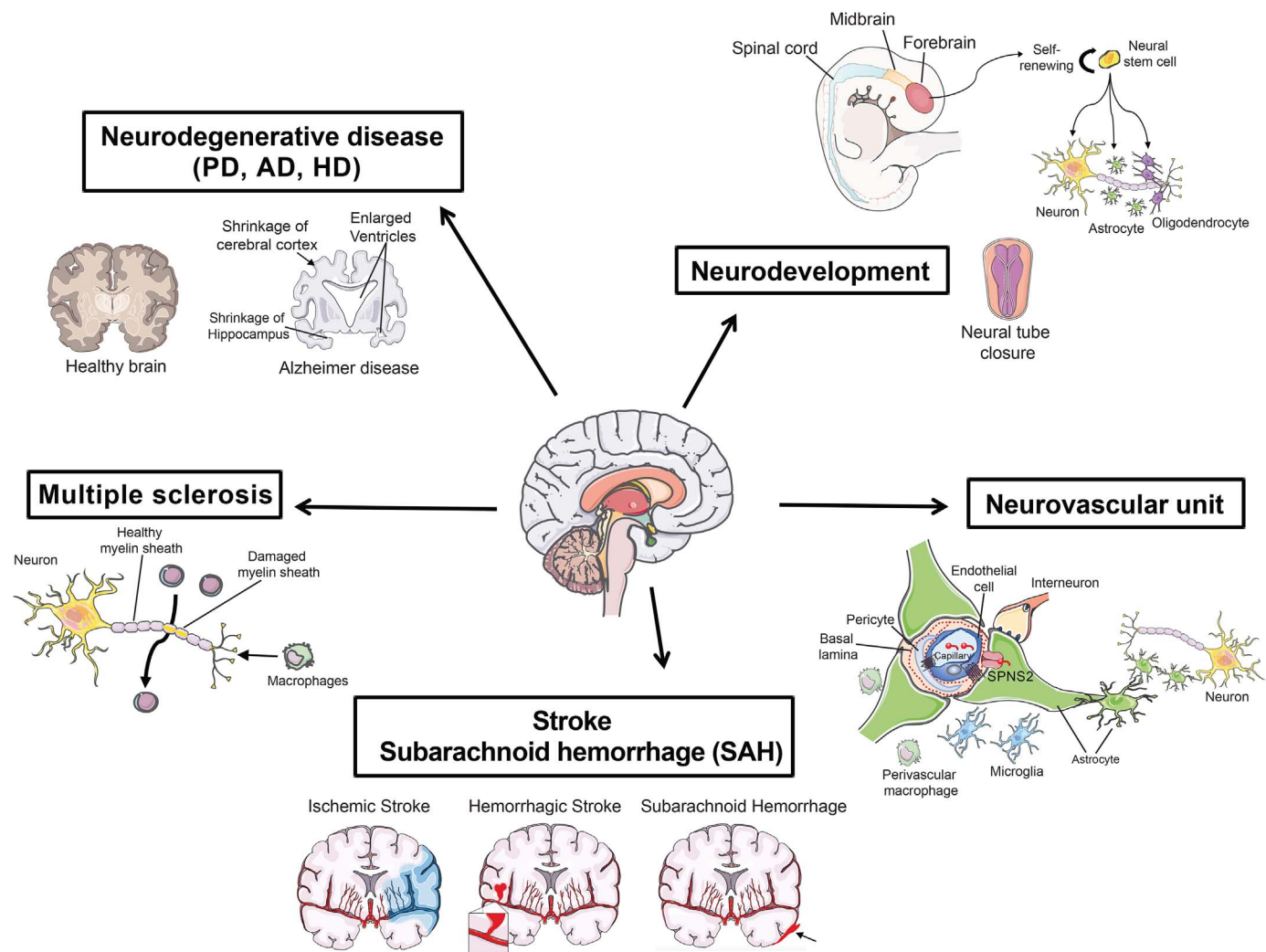


Fig. 5. S1P in the CNS. S1PRs are important for normal CNS development, neural stem cell self-renewal and differentiation, as well as neural tube closure and the formation of the neurovascular unit (NVU). Dysregulation of S1PRs is also implicated in stroke, subarachnoid hemorrhage (SAH), multiple sclerosis, and neurodegenerative diseases such as Parkinson's disease (PD), Alzheimer's disease (AD), and Huntington's disease (HD).

toward pro-healing and anti-inflammatory (M2) macrophages (124, 125). Further, genetic deletion of *Sphk2* in bone marrow in mice is sufficient to protect from folic acid-induced renal fibrosis through the signaling of IFN- γ , CXC ligand 9 (CXCL9), and CXCL10 (126) on infiltrated hematopoietic cells as well as kidney parenchymal cells. Because kidney injury and fibrosis occur in many clinical conditions—including diabetes, hypertension, autoimmune diseases, and infections—the identification of vascular S1PR signaling activation as a renal protective mechanism has revealed numerous therapeutic opportunities such as S1PR1-selective agonists and SPHK inhibitors.

S1P pathway dysregulation is also involved in liver fibrosis, a serious chronic disease that is associated with fatty liver disease, hepatitis, and alcoholism. In human tissues, SPHK1, SPNS2, and S1PR2 expression correlate with the sever-

ity of liver fibrosis (127). Increased liver S1P levels activate hepatic stellate cells, which are located between sinusoidal vessels and hepatocytes, to undergo fibrotic changes (127, 128), and inhibition of S1PR2 with a pharmacological antagonist attenuated liver fibrosis in mice (129). S1P signaling in bone marrow-derived macrophages recruited to the injured liver in mice may be involved in the amplification of fibrotic responses (130, 131). Liver regeneration in mice after partial hepatectomy or bile duct ligation is dependent on S1PR1 expressed by liver sinusoidal endothelial cell (LSEC) activation by ApoM⁺HDL-bound S1P and suppression of Rho signaling. Hepatic sinusoidal vascular remodeling and the restoration of functional liver tissue is impaired in *ApoM*-deficient mice, which showed characteristics of perivascular fibrosis and thrombosis. Conversely, ApoM transgenic mice showed minimal

fibrosis by way of endothelial S1PR1 function in liver sinusoids (40).

Idiopathic pulmonary fibrosis (IPF), a progressive lung disease that occurs in ~0.1% of the population worldwide, leads to disability and death. The S1P pathway is implicated in normal function of the pulmonary vasculature and appears to be impaired in acute lung dysfunction while it is ectopically induced in chronic fibrosis. Patients with IPF have increased S1P levels in serum and bronchoalveolar lavage (BAL) and increased SPHK1 expression in lung alveolar macrophages, which correlated with markers of epithelial mesenchymal transition (EMT) and fibrosis such as α -smooth muscle actin (α -SMA), vimentin, and collagen type 1. In alveolar type II cells in the lung, S1P activates S1PR2 and S1PR3 and induces EMT through the activation of phosphorylated-SMAD3, RhoA-GTP, oxidative stress, and TGF- β 1 release

(132). The use of receptor-selective agonists revealed that although endothelial cell S1PR1 protects from fibrosis and promotes normal lung function, S1PR2 and S1PR3 expressed in fibroblasts and infiltrating myeloid cells promotes fibrosis, such as EMT marker expression, increased proliferation of myofibroblasts, and extracellular matrix deposition (133). The fibrosis-inducing potential of the DNA-damaging agent bleomycin in mouse models was attenuated in *S1pr2*- and *S1pr3*-deficient mice (134, 135). In these preclinical models, collaboration of inflammatory cells and lung parenchymal cells occurs in the complex S1P signaling network to drive fibrotic responses. These observations suggest agonism of vascular endothelial S1PR1 and antagonism of parenchymal S1PR2 and/or S1PR3 as potential therapeutic approaches in fibrotic diseases of multiple organ systems. Whether the results from mouse models will be relevant to human fibrotic disease is not clear and requires large-animal preclinical models and ultimately clinical trials with appropriate S1PR-selective agents.

Therapeutic opportunities

Fingolimod, the first FDA-approved S1PR-targeted drug, is an effective first-line therapy in relapsing remitting multiple sclerosis (RRMS). The primary mechanism of action is as a functional antagonist of S1PR1 expressed on lymphocytes (24), preventing the egress of autoreactive lymphocytes from secondary lymphoid organs and reducing their migration into the CNS, thus slowing disease progression. Substantial adverse events such as first-dose bradycardia (reduction of heart rate) and macular edema (vascular leak into the center of the retina) limits the usage of fingolimod (84). The adverse events can be explained by on-target actions on S1PRs expressed by the heart and retinal vasculature.

Currently, several clinical trials are under way to test the effectiveness of the next-generation S1PR-targeted agents (Table 1) (136). Such agents also induce functional antagonism of S1PR1 in lymphocytes and are more selective in their interaction with other S1PRs. One such compound, siponimod, a selective antagonist of S1PR1 and S1PR5, was shown to have efficacy in both RRMS and progressive forms of MS in human patients but had similar adverse event profiles as fingolimod (137), which confirms that interference with S1PR1 in the heart and vasculature is the cause of these side effects. Siponimod achieved FDA approval in 2019. Siponimod interaction with S1PR5 is thought to enhance survival and/or differentiation of mouse oligodendrocytes, which are defective in MS and neurodegenerative diseases (138). Subsequent studies of fingolimod as well as the recent phase III studies with siponimod (137) found that S1PR

functional antagonist therapy of MS patients protects from accelerated brain atrophy. This finding warrants further studies to examine whether this therapeutic axis is useful in neurodegenerative diseases in general. Furthermore, because S1PR1-dependent autoreactive immune cell trafficking to target organs is common to many autoimmune diseases, this therapeutic approach is likely applicable to other non-CNS autoimmune diseases. As such, clinical trials are under way to treat patients with psoriasis, inflammatory bowel disease, and systemic lupus erythematosus (Table 1).

Biased agonists of S1PR1 may achieve vascular protective effects in diabetes and metabolic and cardiovascular diseases (38). Activation by S1P of S1PR1 in the endothelium leads to receptor down-regulation (endocytosis), which limits the response to agonists. However, biased agonists of endothelial S1PR1 that do not induce S1PR1 endocytosis—and thus, receptor degradation—may induce sustained endothelial protective effects such as NO synthesis, increased barrier function, and endothelial cell survival, among others, thus protecting the vasculature. However, it is possible that such agents could have distinct adverse effect profiles, such as hypotension. Biased agonists for other GPCRs such as μ -opioid receptor, dopamine D2 receptor, and angiotensin receptor-1 are being actively developed to achieve clinical efficacy and reduced adverse effect profiles (139). In some cases, such as in the case of the μ -opioid receptor, a biased agonist potentially achieved analgesia with limited adverse effects (140).

Small-molecule pharmacological antagonists have been used to target other S1PRs in preclinical models of inflammation and fibrosis (129, 131, 135, 141). This approach may potentially lead to adverse events such as hearing defects (35), lymphoid neoplasms (87), and epileptic seizures (142), which have been observed in *S1pr2* genetic loss of function mouse models. S1PR redundancy may also lead to limited efficacy in targeting a single receptor isoform, and therefore, targeting multiple receptor isoforms that carry out similar biological effects may prove to be more effective.

S1P chaperones allow an additional opportunity for S1PR modulation. For example, therapeutic delivery of engineered ApoM—namely, ApoM-Fc—suppresses inflammation (39, 41) and promotes vascular homeostasis in mouse models of myocardial ischemia-reperfusion injury, chronic hypertension, and experimental stroke (67). In sharp contrast to small-molecule S1PR1-targeted drugs, ApoM-S1P treatment did not lead to lymphopenia, suggesting the inability of chaperone-bound S1P to down-regulate S1PR1 in secondary lymphoid organs. This ApoM-S1P therapeutic approach may selectively activate endothelial cell S1PRs without inducing functional antagonism of im-

mune cell S1PRs. However, ApoM-S1P can activate multiple receptors and therefore could potentially lead to undesirable side effects by activating pro-inflammatory S1PR2 and S1PR3. Selective activation of S1PRs by modified chaperone-based approaches may need to be developed to provide therapeutic utility.

Selective activation of S1P transporters may have utility in cancer metastasis and anemia (34, 96). In metastasis, SPNS2 activation, which allows S1P export and signaling by immune cell S1PRs to regulate trafficking, may enhance CD8⁺ T cell and NK cell infiltration in tumors and potentially enhance antitumor immunity. In anemia, activation of the MFSD2B transporter in RBCs may help reduce the concentration of sphingosine and S1P in the cell membranes of RBCs, which was shown to reduce recovery from anemia in mice.

The inhibition of the metabolic enzymes involved in S1P biosynthesis may also have therapeutic utility. An S1P lyase inhibitor induced lymphopenia and suppressed autoimmune inflammation by enhancing S1P levels (Table 1) (143). However, broad-spectrum sphingolipid alterations induced by inhibition of S1P lyase could lead to side effects. Similarly, even though SPHK inhibitors have been tested in preclinical models to inhibit the overactive SPHK1 enzyme in cancer (144), recent work with highly specific and potent inhibitors suggest that they have limited utility, indicating that off-target effects of early SPHK-targeted agents may be involved (145). SPHK2 inhibitors led to increased plasma S1P and attenuation of inflammatory and fibrotic responses (146). Because modulation of S1P metabolic enzymes led to broad changes in the metabolite concentrations and flux of not only sphingolipids but also other phospholipids, this approach will need to carefully consider lipid metabolism before clinical application.

Conclusions and future perspectives

The use of mouse models and modern pharmacological tools has led to our current understanding of the biology of the S1P-S1PR signaling system. With the application of S1PR-targeted drugs in humans, much has been learned about the role of this pathway in the immune system and its potential future application in immunological, neurological, and perhaps cardiovascular and fibrotic diseases. S1PR1 functional antagonists that induce irreversible S1PR internalization and degradation are the most advanced class of compounds, whereas competitive antagonists have not yet been tested in clinical trials. Perhaps functional antagonism of the receptors may be preferable to block S1PRs in ligand-rich compartments and may improve efficacy in the clinic. However, the emerging biological understanding and clinical research suggest additional opportunities for therapeutic application, including transporter modulators, biased agonists of

Table 1. S1P modulators under development or in clinical testing. RRMS, relapsing remitting multiple sclerosis; PPMS, primary progressive multiple sclerosis; SPMS, secondary progressive multiple sclerosis; RMS, relapsing multiple sclerosis; UC, ulcerative colitis; CIPD, chronic inflammatory demyelinating polyradiculoneuropathy; ALS, amyotrophic lateral sclerosis; SLE, systemic lupus erythematosus; SCLE, subacute cutaneous lupus erythematosus; RA, rheumatoid arthritis; AMD, age-related macular degeneration.

Molecule name	Drug name	Targeted protein	Sponsor	Proposed or approved use	Current trial phase	CT ID
BAF312	Fingolimod (Gilenya)	S1PR1, -3, -4, and -5	Novartis	RRMS	Marketed	
				Transplant	III	NCT00099801
				CIPD	III	NCT01625182
				PPMS	III	NCT00731692 (INFORMS)
				ALS	II	NCT01786174
	Siponimod (Mayzent)	S1PR1 and -5	Novartis	Asthma	II	NCT00785083
				SPMS	FDA approved	NCT01665144 (EXPAND)
				RRMS	II	NCT01185821
				Active dermatomyositis	II	NCT02029274
				Hemorrhagic stroke, ICH	II	NCT03338998
KRP-203		S1PR1 and -5	Novartis	Hepatic impairment	I	NCT01565902
				UC	II	NCT01375179
				SCLE	II	NCT01294774
				Hematological malignancies	I	NCT01830010
ACT-128800	Ponesimod	S1PR1, -3, and -5	Actelion	MS	III	NCT03232073 (OPTIMUM-LT)
ACT-334441	Cerenimod	S1PR1 and -5	Actelion	Psoriasis	II	NCT01208090
APD334	Etrasimod	S1PR1	Arena Pharmaceuticals	SLE	II	NCT02472795
RPC1063	Ozanimod	S1PR1 and -5	Receptos (Celgene)	UC	III	NCT03945188
				Crohn	III	NCT03440372
				UC	III	NCT03440385
				RMS	III	NCT02435992
ONO-4641	Ceralifimod	S1PR1 and -5	Ono	RRMS	II	NCT02531126
MT-1303	Amiselimod	S1PR1, -4, and -5	Mitsubishi Tanabe	RRMS	II	NCT01081782 (DreaMS)
				Crohn	II	NCT01742052
				Psoriasis	II	NCT02389790
				SLE	I	NCT01987843
AKP11		S1PR1	Akaal Pharma	Atopic dermatitis	II	NCT02307643
GSK-2018682		S1PR1	GlaxoSmithKline	RA	II	ACTRN12617000763347
BMS-986104		S1PR1	GlaxoSmithKline	RRMS	I	ACTRN12617001223325
CS-0777		S1PR1	GlaxoSmithKline	RRMS	I	NCT01431937 (PIA114347)
LT1009 (Sonepcizumab)	ASOPNEP	S1P antibody	Lpath Inc / Pfizer	RA	I	NCT02211469
	iSONEP			MS	I	NCT00616733
				Renal cell carcinoma	II	NCT01762033
ABC294640	Opaganib (Yeliva®)	SphK2	RedHill Biopharma Limited	AMD	II	NCT01414153 (Nexus)
				Cholangiocarcinoma	II	NCT03377179
				Hepatocellular carcinoma	II	NCT02939807
DHS (L-threo-dihydrosphingosine)	Safingol	SphK1	MSKCC/NCI	Solid Tumor	I	NCT01488513 (ABC-101)
LX3305		SPGL1	Lexicon Pharmaceuticals	Locally advanced or metastatic solid tumors	1	NCT00084812
				RA	II	NCT00903383

S1PRs, chaperone-based strategies, and ligand neutralization approaches. Such approaches, which attenuate abnormal S1PR signaling and normalize spatial SIP gradients, may be useful to combat diseases in the context of the complex SIP signaling mechanisms that are prevalent in many organ systems.

REFERENCES AND NOTES

1. T. Hla, Genomic insights into mediator lipidomics. *Prostaglandins Other Lipid Mediat.* **77**, 197–209 (2005). doi: [10.1016/j.prostaglandins.2005.06.008](#); pmid: [16099404](#)
2. T. Shimizu, Lipid mediators in health and disease: Enzymes and receptors as therapeutic targets for the regulation of immunity and inflammation. *Annu. Rev. Pharmacol. Toxicol.* **49**, 123–150 (2009). doi: [10.1146/annurev.pharmtox.011008.145616](#); pmid: [18834304](#)
3. R. L. Proia, T. Hla, Emerging biology of sphingosine-1-phosphate: Its role in pathogenesis and therapy. *J. Clin. Invest.* **125**, 1379–1387 (2015). doi: [10.1172/JCI76369](#); pmid: [25831442](#)
4. W. Stoffel, G. Sticht, D. LeKim, Metabolism of sphingosine bases. IX. Degradation in vitro of dihydrosphingosine and dihydrosphingosine phosphate to palmitaldehyde and ethanolamine phosphate. *Hoppe Seyler's Z. Physiol. Chem.* **349**, 1745–1748 (1968). doi: [10.1515/bchm2.1968.349.2.1745](#); pmid: [5707041](#)
5. T. K. Ghosh, J. Bian, D. L. Gill, Intracellular calcium release mediated by sphingosine derivatives generated in cells. *Science* **248**, 1653–1656 (1990). doi: [10.1126/science.2163543](#); pmid: [2163543](#)
6. H. Zhang et al., Sphingosine-1-phosphate, a novel lipid, involved in cellular proliferation. *J. Cell Biol.* **114**, 155–167 (1991). doi: [10.1083/jcb.114.1.155](#); pmid: [2050740](#)
7. J. Chun, T. Hla, K. R. Lynch, S. Spiegel, W. H. Moolenaar, International Union of Basic and Clinical Pharmacology. LXXVIII. Lysophospholipid receptor nomenclature. *Pharmacol. Rev.* **62**, 579–587 (2010). doi: [10.1124/pr.110.003111](#); pmid: [21079037](#)
8. S. E. Alvarez et al., Sphingosine-1-phosphate is a missing cofactor for the E3 ubiquitin ligase TRAF2. *Nature* **465**, 1084–1088 (2010). doi: [10.1038/nature09128](#); pmid: [20577214](#)
9. N. C. Hait et al., Regulation of histone acetylation in the nucleus by sphingosine-1-phosphate. *Science* **325**, 1254–1257 (2009). doi: [10.1126/science.1176709](#); pmid: [19729656](#)
10. G. M. Strub et al., Sphingosine-1-phosphate produced by sphingosine kinase 2 in mitochondria interacts with prohibitin 2 to regulate complex IV assembly and respiration. *FASEB J.* **25**, 600–612 (2011). doi: [10.1096/fj.10.167502](#); pmid: [20959514](#)
11. T. Kajimoto et al., Activation of atypical protein kinase C by sphingosine 1-phosphate revealed by an aPKC-specific activity reporter. *Sci. Signal.* **12**, eaat6662 (2019). doi: [10.1126/scisignal.aat6662](#); pmid: [30600259](#)
12. S. Pannier Selvam et al., Binding of the sphingolipid S1P to hTERT stabilizes telomerase at the nuclear periphery by allosterically mimicking protein phosphorylation. *Sci. Signal.* **8**, ra58 (2015). doi: [10.1126/scisignal.aaa4998](#); pmid: [26082434](#)
13. N. Etemadi et al., TRAF2 regulates TNF and NF- κ B signalling to suppress apoptosis and skin inflammation independently of Sphingosine kinase 1. *eLife* **4**, e10592 (2015). doi: [10.7554/eLife.10592](#); pmid: [26701909](#)
14. K. Mendelson, Y. Lan, T. Hla, T. Evans, Maternal or zygotic sphingosine kinase is required to regulate zebrafish cardiogenesis. *Dev. Dyn.* **244**, 948–954 (2015). doi: [10.1002/dvdy.24288](#); pmid: [25997406](#)
15. K. Mendelson et al., The ceramide synthase 2b gene mediates genomic sensing and regulation of sphingosine levels during zebrafish embryogenesis. *eLife* **6**, e21992 (2017). doi: [10.7554/eLife.21992](#); pmid: [28956531](#)
16. Y. Xiong et al., Sphingosine kinases are not required for inflammatory responses in macrophages. *J. Biol. Chem.* **288**, 32563–32573 (2013). doi: [10.1074/jbc.M113.483750](#); pmid: [24081141](#)
17. Y. A. Hannun, L. M. Obeid, Principles of bioactive lipid signalling: Lessons from sphingolipids. *Nat. Rev. Mol. Cell Biol.* **9**, 139–150 (2008). doi: [10.1038/nrm2329](#); pmid: [18216770](#)
18. N. Murata et al., Interaction of sphingosine 1-phosphate with plasma components, including lipoproteins, regulates the lipid receptor-mediated actions. *Biochem. J.* **352**, 809–815 (2000). doi: [10.1042/bj3520809](#); pmid: [11104690](#)
19. S. R. Schwab et al., Lymphocyte sequestration through S1P lyase inhibition and disruption of S1P gradients. *Science* **309**, 1735–1739 (2005). doi: [10.1126/science.1113640](#); pmid: [16151014](#)
20. C. Christoffersen et al., Endothelium-protective sphingosine-1-phosphate provided by HDL-associated apolipoprotein M. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9613–9618 (2011). doi: [10.1073/pnas.1103187108](#); pmid: [21606363](#)
21. K. Yanagida, T. Hla, Vascular and Immunobiology of the Circulatory Sphingosine 1-Phosphate Gradient. *Annu. Rev. Physiol.* **79**, 67–91 (2017). doi: [10.1146/annurev-physiol-021014-071635](#); pmid: [27813829](#)
22. S. Lovric et al., Mutations in sphingosine-1-phosphate lyase cause nephrosis with ichthyosis and adrenal insufficiency. *J. Clin. Invest.* **127**, 912–928 (2017). doi: [10.1172/JCI89626](#); pmid: [28165339](#)
23. R. L. Santos-Cortez et al., Autosomal-recessive hearing impairment due to rare missense variants within S1PR2. *Am. J. Hum. Genet.* **98**, 331–338 (2016). doi: [10.1016/j.ajhg.2015.12.004](#); pmid: [26805784](#)
24. V. Brinkmann et al., Fingolimod (FTY720): Discovery and development of an oral drug to treat multiple sclerosis. *Nat. Rev. Drug Discov.* **9**, 883–897 (2010). doi: [10.1038/nrd3248](#); pmid: [21031003](#)
25. A. D. Lander, Morpheus unbound: Reimagining the morphogen gradient. *Cell* **128**, 245–256 (2007). doi: [10.1016/j.cell.2007.01.004](#); pmid: [17254964](#)
26. M. Maceyka, S. Spiegel, Sphingolipid metabolites in inflammatory disease. *Nature* **510**, 58–67 (2014). doi: [10.1038/nature13475](#); pmid: [24899305](#)
27. A. Kawahara et al., The sphingolipid transporter spns2 functions in migration of zebrafish myocardial precursors. *Science* **323**, 524–527 (2009). doi: [10.1126/science.1167449](#); pmid: [19074308](#)
28. E. Kupperman, S. An, N. Osborne, S. Waldron, D. Y. Stainier, A sphingosine-1-phosphate receptor regulates cell migration during vertebrate heart development. *Nature* **406**, 192–195 (2000). doi: [10.1038/35018092](#); pmid: [10910360](#)
29. N. Osborne et al., The spinster homolog, two of hearts, is required for sphingosine 1-phosphate signaling in zebrafish. *Curr. Biol.* **18**, 1882–1888 (2008). doi: [10.1016/j.cub.2008.10.061](#); pmid: [19062281](#)
30. S. Fukuhara et al., The sphingosine-1-phosphate transporter Spns2 expressed on endothelial cells regulates lymphocyte trafficking in mice. *J. Clin. Invest.* **122**, 1416–1426 (2012). doi: [10.1172/JCI60746](#); pmid: [22406534](#)
31. A. Mendoza et al., The transporter Spns2 is required for secretion of lymph but not plasma sphingosine-1-phosphate. *Cell Rep.* **2**, 1104–1110 (2012). doi: [10.1016/j.celrep.2012.09.021](#); pmid: [23103166](#)
32. W. D. Ramos-Perez, V. Fang, D. Escalante-Alcalde, M. Cammer, S. R. Schwab, A map of the distribution of sphingosine 1-phosphate in the spleen. *Nat. Immunol.* **16**, 1245–1252 (2015). doi: [10.1038/ni.3296](#); pmid: [26502404](#)
33. V. Fang et al., Gradients of the signaling lipid S1P in lymph nodes position natural killer cells and regulate their interferon- γ response. *Nat. Immunol.* **18**, 15–25 (2017). doi: [10.1038/ni.3619](#); pmid: [27841869](#)
34. T. M. Vu et al., Mfsd2b is essential for the sphingosine-1-phosphate export in erythrocytes and platelets. *Nature* **550**, 524–528 (2017). doi: [10.1038/nature24053](#); pmid: [29045386](#)
35. J. Chen et al., Spinster homolog 2 (spns2) deficiency causes early onset progressive hearing loss. *PLOS Genet.* **10**, e1004688 (2014). doi: [10.1371/journal.pgen.1004688](#); pmid: [25356849](#)
36. C. Fang et al., S1P transporter SPNS2 regulates proper postnatal retinal morphogenesis. *FASEB J.* **32**, 3597–3613 (2018). doi: [10.1096/fj.201701116R](#); pmid: [29452570](#)
37. J. K. Fleming, J. M. Wojciak, Measuring Sphingosine-1-phosphate/protein interactions with the kinetic exclusion assay. *Methods Mol. Biol.* **1697**, 1–8 (2018). pmid: [28349502](#)
38. S. Galvani et al., HDL-bound sphingosine 1-phosphate acts as a biased agonist for the endothelial cell receptor S1P1 to limit vascular inflammation. *Sci. Signal.* **8**, ra79 (2015). doi: [10.1126/scisignal.aaa2581](#); pmid: [26268607](#)
39. P. M. Christensen et al., Impaired endothelial barrier function in apolipoprotein M-deficient mice is dependent on sphingosine-1-phosphate receptor 1. *FASEB J.* **30**, 2351–2359 (2016). doi: [10.1096/fj.201500064](#); pmid: [26956418](#)
40. B. S. Ding et al., HDL activation of endothelial sphingosine-1-phosphate receptor-1 (S1P₁) promotes regeneration and suppresses fibrosis in the liver. *JCI Insight* **1**, e87058 (2016). doi: [10.1172/jci.insight.87058](#); pmid: [28018969](#)
41. M. Kurano et al., Apolipoprotein M protects lipopolysaccharide-treated mice from death and organ injury. *Thromb. Haemost.* **118**, 1021–1035 (2018). doi: [10.1055/s-0038-1641750](#); pmid: [29669385](#)
42. V. A. Blaho et al., HDL-bound sphingosine-1-phosphate restrains lymphopoiesis and neuroinflammation. *Nature* **523**, 342–346 (2015). doi: [10.1038/nature14462](#); pmid: [26053123](#)
43. M. H. Lee et al., S1P in HDL promotes interaction between SR-BI and S1PR1 and activates S1PR1-mediated biological functions: Calcium flux and S1PR1 internalization. *J. Lipid Res.* **58**, 325–338 (2017). doi: [10.1194/jlr.M070706](#); pmid: [27881715](#)
44. K. Venkataraman et al., Vascular endothelium as a contributor of plasma sphingosine 1-phosphate. *Circ. Res.* **102**, 669–676 (2008). doi: [10.1161/CIRCRESAHA.107.165845](#); pmid: [18258856](#)
45. T. I. Arnon et al., GRK2-dependent S1PR1 desensitization is required for lymphocytes to overcome their attraction to blood. *Science* **333**, 1898–1903 (2011). doi: [10.1126/science.1208248](#); pmid: [21960637](#)
46. S. Thangada et al., Cell-surface residence of sphingosine 1-phosphate receptor 1 on lymphocytes determines lymphocyte egress kinetics. *J. Exp. Med.* **207**, 1475–1483 (2010). doi: [10.1084/jem.20091343](#); pmid: [20584883](#)
47. T. Willinger, S. M. Ferguson, J. P. Pereira, P. De Camilli, R. A. Flavell, Dynamins 2-dependent endocytosis is required for sustained S1PR1 signaling. *J. Exp. Med.* **211**, 685–700 (2014). doi: [10.1084/jem.20131343](#); pmid: [24638168](#)
48. A. Nomachi et al., Moesin controls clathrin-mediated S1PR1 internalization in T cells. *PLOS ONE* **8**, e82590 (2013). doi: [10.1371/journal.pone.0082590](#); pmid: [24358210](#)
49. M. L. Oo et al., Engagement of S1P₁-degradative mechanisms leads to vascular leak in mice. *J. Clin. Invest.* **121**, 2290–2300 (2011). doi: [10.1172/JCI45403](#); pmid: [21555855](#)
50. M. Kono et al., Sphingosine-1-phosphate receptor 1 reporter mice reveal receptor activation sites in vivo. *J. Clin. Invest.* **124**, 2076–2086 (2014). doi: [10.1172/JCI71194](#); pmid: [24667638](#)
51. M. Kono et al., Bioluminescence imaging of G protein-coupled receptor activation in living mice. *Nat. Commun.* **8**, 1163 (2017). doi: [10.1038/s41467-017-01340-7](#); pmid: [29079828](#)
52. E. Studer et al., Conjugated bile acids activate the sphingosine-1-phosphate receptor 2 in primary rodent hepatocytes. *Hepatology* **55**, 267–276 (2012). doi: [10.1002/hep.24681](#); pmid: [21932398](#)
53. C. Feistritzer, M. Riewald, Endothelial barrier protection by activated protein C through PAR1-dependent sphingosine 1-phosphate receptor-1 crossactivation. *Blood* **105**, 3178–3184 (2005). doi: [10.1182/blood-2004-10-3985](#); pmid: [15626732](#)
54. P. A. Singleton, S. M. Dudek, S. F. Ma, J. G. Garcia, Transactivation of sphingosine 1-phosphate receptors is essential for vascular barrier regulation. Novel role for hyaluronan and CD44 receptor family. *J. Biol. Chem.* **281**, 34381–34393 (2006). doi: [10.1074/jbc.M603680200](#); pmid: [16963454](#)
55. L. R. Shiow et al., CD69 acts downstream of interferon- α/β to inhibit S1P₁ and lymphocyte egress from lymphoid organs. *Nature* **440**, 540–544 (2006). doi: [10.1038/nature04606](#); pmid: [16525420](#)
56. A. Kempf et al., The sphingolipid receptor S1PR2 is a receptor for Nogo-a repressing synaptic plasticity. *PLOS Biol.* **12**, e1001763 (2014). doi: [10.1371/journal.pbio.1001763](#); pmid: [24453941](#)
57. Y. Hisano et al., Lysolipid receptor cross-talk regulates lymphatic endothelial junctions in lymph nodes. *J. Exp. Med.* **216**, 1582–1598 (2019). doi: [10.1084/jem.20181895](#); pmid: [31147448](#)
58. B. J. Laidlaw, E. E. Gray, Y. Zhang, F. Ramirez-Valle, J. G. Cyster, Sphingosine-1-phosphate receptor 2 restrains egress of $\gamma\delta$ T cells from the skin. *J. Exp. Med.* **216**, 1487–1496 (2019). doi: [10.1084/jem.20190114](#); pmid: [31160320](#)
59. M. A. Hanson et al., Crystal structure of a lipid G protein-coupled receptor. *Science* **335**, 851–855 (2012). doi: [10.1126/science.1215904](#); pmid: [22444443](#)
60. K. Gaengel, G. Genové, A. Armulik, C. Betsholtz, Endothelial-mural cell signaling in vascular development

- and angiogenesis. *Arterioscler. Thromb. Vasc. Biol.* **29**, 630–638 (2009). doi: [10.1161/ATVBAHA.107.165121](https://doi.org/10.1161/ATVBAHA.107.165121); pmid: [19164813](https://pubmed.ncbi.nlm.nih.gov/19164813/)
61. B. Jung *et al.*, Flow-regulated endothelial S1P receptor-1 signaling sustains vascular development. *Dev. Cell* **23**, 600–610 (2012). doi: [10.1016/j.devcel.2012.07.015](https://doi.org/10.1016/j.devcel.2012.07.015); pmid: [22975328](https://pubmed.ncbi.nlm.nih.gov/22975328/)
 62. M. J. Lee *et al.*, Vascular endothelial cell adherens junction assembly and morphogenesis induced by sphingosine-1-phosphate. *Cell* **99**, 301–312 (1999). doi: [10.1016/S0092-8674\(00\)81661-X](https://doi.org/10.1016/S0092-8674(00)81661-X); pmid: [10555146](https://pubmed.ncbi.nlm.nih.gov/10555146/)
 63. J. H. Paik *et al.*, Sphingosine 1-phosphate receptor regulation of N-cadherin mediates vascular stabilization. *Genes Dev.* **18**, 2392–2403 (2004). doi: [10.1101/gad.1227804](https://doi.org/10.1101/gad.1227804); pmid: [15371328](https://pubmed.ncbi.nlm.nih.gov/15371328/)
 64. Y. Xiong, P. Yang, R. L. Proia, T. Hla, Erythrocyte-derived sphingosine 1-phosphate is essential for vascular development. *J. Clin. Invest.* **124**, 4823–4828 (2014). doi: [10.1172/JCI77685](https://doi.org/10.1172/JCI77685); pmid: [25250575](https://pubmed.ncbi.nlm.nih.gov/25250575/)
 65. S. L. Gazit *et al.*, Platelet and erythrocyte sources of S1P are redundant for vascular development and homeostasis, but both rendered essential after plasma S1P depletion in anaphylactic shock. *Circ. Res.* **119**, e110–e126 (2016). doi: [10.1161/CIRCRESAHA.116.308929](https://doi.org/10.1161/CIRCRESAHA.116.308929); pmid: [27582371](https://pubmed.ncbi.nlm.nih.gov/27582371/)
 66. N. Burg, S. Swendeman, S. Worgall, T. Hla, J. E. Salmon, Sphingosine 1-phosphate receptor 1 signaling maintains endothelial cell barrier function and protects against immune complex-induced vascular injury. *Arthritis Rheumatol.* **70**, 1879–1889 (2018). doi: [10.1002/art.40558](https://doi.org/10.1002/art.40558); pmid: [29781582](https://pubmed.ncbi.nlm.nih.gov/29781582/)
 67. S. L. Swendeman *et al.*, An engineered S1P chaperone attenuates hypertension and ischemic injury. *Sci. Signal.* **10**, eaal2722 (2017). doi: [10.1126/scisignal.aal2722](https://doi.org/10.1126/scisignal.aal2722); pmid: [28811382](https://pubmed.ncbi.nlm.nih.gov/28811382/)
 68. L. Perségol *et al.*, Small dense HDLs display potent vasorelaxing activity, reflecting their elevated content of sphingosine-1-phosphate. *J. Lipid Res.* **59**, 25–34 (2018). doi: [10.1194/jlr.M076927](https://doi.org/10.1194/jlr.M076927); pmid: [29150495](https://pubmed.ncbi.nlm.nih.gov/29150495/)
 69. Y. Zeng, R. H. Adamson, F. R. Curry, J. M. Tarbell, Sphingosine-1-phosphate protects endothelial glycocalyx by inhibiting syndecan-1 shedding. *Am. J. Physiol. Heart Circ. Physiol.* **306**, H363–H372 (2014). doi: [10.1152/ajpheart.00687.2013](https://doi.org/10.1152/ajpheart.00687.2013); pmid: [24285115](https://pubmed.ncbi.nlm.nih.gov/24285115/)
 70. A. Cantalupo *et al.*, Nogo-B regulates endothelial sphingolipid homeostasis to control vascular function and blood pressure. *Nat. Med.* **21**, 1028–1037 (2015). doi: [10.1038/nm.3934](https://doi.org/10.1038/nm.3934); pmid: [26301690](https://pubmed.ncbi.nlm.nih.gov/26301690/)
 71. A. Cantalupo *et al.*, S1PR1 (sphingosine-1-phosphate receptor 1) signaling regulates blood flow and pressure. *Hypertension* **70**, 426–434 (2017). doi: [10.1161/HYPERTENSIONAHA.117.09088](https://doi.org/10.1161/HYPERTENSIONAHA.117.09088); pmid: [28607130](https://pubmed.ncbi.nlm.nih.gov/28607130/)
 72. M. Książek, U. Baranowska, A. Chabowski, M. Baranowski, Arteriovenous sphingosine-1-phosphate differences across selected organs of the rat. *Cell. Physiol. Biochem.* **54**, 67–77 (2018). doi: [10.1159/000486223](https://doi.org/10.1159/000486223); pmid: [29316552](https://pubmed.ncbi.nlm.nih.gov/29316552/)
 73. A. Skoura *et al.*, Essential role of sphingosine 1-phosphate receptor 2 in pathological angiogenesis of the mouse retina. *J. Clin. Invest.* **117**, 2506–2516 (2007). doi: [10.1172/JCI31123](https://doi.org/10.1172/JCI31123); pmid: [17710232](https://pubmed.ncbi.nlm.nih.gov/17710232/)
 74. S. Caballero *et al.*, Anti-sphingosine-1-phosphate monoclonal antibodies inhibit angiogenesis and sub-retinal fibrosis in a murine model of laser-induced choroidal neovascularization. *Exp. Eye Res.* **88**, 367–377 (2009). doi: [10.1016/j.exer.2008.07.012](https://doi.org/10.1016/j.exer.2008.07.012); pmid: [18723015](https://pubmed.ncbi.nlm.nih.gov/18723015/)
 75. M. V. Simón, F. H. Prado Spalm, L. E. Politi, N. P. Rotstein, Sphingosine-1-phosphate is a crucial signal for migration of retina müller glial cells. *Invest. Ophthalmol. Vis. Sci.* **56**, 5808–5815 (2015). doi: [10.1167/iov.14-16195](https://doi.org/10.1167/iov.14-16195); pmid: [26325420](https://pubmed.ncbi.nlm.nih.gov/26325420/)
 76. H. Fukui *et al.*, S1P-Yap1 signaling regulates endoderm formation required for cardiac precursor cell migration in zebrafish. *Dev. Cell* **31**, 128–136 (2014). doi: [10.1016/j.devcel.2014.08.014](https://doi.org/10.1016/j.devcel.2014.08.014); pmid: [25313964](https://pubmed.ncbi.nlm.nih.gov/25313964/)
 77. H. Clay *et al.*, Sphingosine 1-phosphate receptor-1 in cardiomyocytes is required for normal cardiac development. *Dev. Biol.* **418**, 157–165 (2016). doi: [10.1016/j.ydbio.2016.06.024](https://doi.org/10.1016/j.ydbio.2016.06.024); pmid: [27333774](https://pubmed.ncbi.nlm.nih.gov/27333774/)
 78. A. Sharma *et al.*, Stage-specific effects of bioactive lipids on human iPSC cardiac differentiation and cardiomyocyte proliferation. *Sci. Rep.* **8**, 6618 (2018). doi: [10.1038/s41598-018-24954-3](https://doi.org/10.1038/s41598-018-24954-3); pmid: [29700394](https://pubmed.ncbi.nlm.nih.gov/29700394/)
 79. R. Ochi, Y. Momose, K. Oyama, W. R. Giles, Sphingosine-1-phosphate effects on guinea pig atrial myocytes: Alterations in action potentials and K⁺ currents. *Cardiovasc. Res.* **70**, 88–96 (2006). doi: [10.1016/j.cardiores.2006.01.010](https://doi.org/10.1016/j.cardiores.2006.01.010); pmid: [16545787](https://pubmed.ncbi.nlm.nih.gov/16545787/)
 80. B. Levkau *et al.*, High-density lipoprotein stimulates myocardial perfusion in vivo. *Circulation* **110**, 3355–3359 (2004). doi: [10.1161/01.CIR.0000147827.43912.AE](https://doi.org/10.1161/01.CIR.0000147827.43912.AE); pmid: [15545521](https://pubmed.ncbi.nlm.nih.gov/15545521/)
 81. D. A. Vessey *et al.*, Role of sphingosine kinase activity in protection of heart against ischemia reperfusion injury. *Med. Sci. Monit.* **12**, BR318–BR324 (2006). pmid: [17006394](https://pubmed.ncbi.nlm.nih.gov/17006394/)
 82. C. G. Santos-Gallego *et al.*, Sphingosine-1-phosphate receptor agonist fingolimod increases myocardial salvage and decreases adverse postinfarction left ventricular remodeling in a porcine model of ischemia/reperfusion. *Circulation* **133**, 954–966 (2016). doi: [10.1161/CIRCULATIONAHA.115.012427](https://doi.org/10.1161/CIRCULATIONAHA.115.012427); pmid: [26826180](https://pubmed.ncbi.nlm.nih.gov/26826180/)
 83. Y. Zhang *et al.*, Endothelial Nogo-B regulates sphingolipid biosynthesis to promote pathological cardiac hypertrophy during chronic pressure overload. *JCI Insight* **1**, e85484 (2016). doi: [10.1172/jci.insight.85484](https://doi.org/10.1172/jci.insight.85484); pmid: [27158676](https://pubmed.ncbi.nlm.nih.gov/27158676/)
 84. J. Camm, T. Hla, R. Bakshi, V. Brinkmann, Cardiac and vascular effects of fingolimod: Mechanistic basis and clinical implications. *Am. Heart J.* **168**, 632–644 (2014). doi: [10.1016/j.ahj.2014.06.028](https://doi.org/10.1016/j.ahj.2014.06.028); pmid: [25440790](https://pubmed.ncbi.nlm.nih.gov/25440790/)
 85. T. I. Arnon, R. M. Horton, I. L. Grigorova, J. G. Cyster, Visualization of splenic marginal zone B-cell shuttling and follicular B-cell egress. *Nature* **493**, 684–688 (2013). doi: [10.1038/nature11738](https://doi.org/10.1038/nature11738); pmid: [23263181](https://pubmed.ncbi.nlm.nih.gov/23263181/)
 86. K. Tedford *et al.*, The opposing forces of shear flow and sphingosine-1-phosphate control marginal zone B cell shuttling. *Nat. Commun.* **8**, 2261 (2017). doi: [10.1038/s41467-017-02482-4](https://doi.org/10.1038/s41467-017-02482-4); pmid: [29273735](https://pubmed.ncbi.nlm.nih.gov/29273735/)
 87. J. R. Muppidi *et al.*, Loss of signalling via Gα13 in germinal centre B-cell-derived lymphoma. *Nature* **516**, 254–258 (2014). doi: [10.1038/nature13765](https://doi.org/10.1038/nature13765); pmid: [25274307](https://pubmed.ncbi.nlm.nih.gov/25274307/)
 88. P. Chongsathidkiet *et al.*, Sequestration of T cells in bone marrow in the setting of glioblastoma and other intracranial tumors. *Nat. Med.* **24**, 1459–1468 (2018). doi: [10.1038/s41591-018-0135-2](https://doi.org/10.1038/s41591-018-0135-2); pmid: [30104766](https://pubmed.ncbi.nlm.nih.gov/30104766/)
 89. C. N. Skon *et al.*, Transcriptional downregulation of S1pr1 is required for the establishment of resident memory CD8⁺ T cells. *Nat. Immunol.* **14**, 1285–1293 (2013). doi: [10.1038/ni.2745](https://doi.org/10.1038/ni.2745); pmid: [24162775](https://pubmed.ncbi.nlm.nih.gov/24162775/)
 90. L. K. Mackay *et al.*, Cutting edge: CD69 interference with sphingosine-1-phosphate receptor function regulates peripheral T cell retention. *J. Immunol.* **194**, 2059–2063 (2015). doi: [10.1049/jimmunol.1402256](https://doi.org/10.1049/jimmunol.1402256); pmid: [25624457](https://pubmed.ncbi.nlm.nih.gov/25624457/)
 91. A. Drouillard *et al.*, Human naive and memory T cells display opposite migratory responses to sphingosine-1 phosphate. *J. Immunol.* **200**, 551–557 (2018). doi: [10.1049/jimmunol.1701278](https://doi.org/10.1049/jimmunol.1701278); pmid: [29237776](https://pubmed.ncbi.nlm.nih.gov/29237776/)
 92. A. Drouillard *et al.*, S1PR5 is essential for human natural killer cell migration toward sphingosine-1 phosphate. *J. Allergy Clin. Immunol.* **141**, 2265–2268.e1 (2018). doi: [10.1016/j.jaci.2017.11.022](https://doi.org/10.1016/j.jaci.2017.11.022); pmid: [29248494](https://pubmed.ncbi.nlm.nih.gov/29248494/)
 93. J. R. Teijaro *et al.*, S1PR1-mediated IFNAR1 degradation modulates plasmacytoid dendritic cell interferon-α autoamplification. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1351–1356 (2016). doi: [10.1073/pnas.1525356113](https://doi.org/10.1073/pnas.1525356113); pmid: [26787880](https://pubmed.ncbi.nlm.nih.gov/26787880/)
 94. Y. Huang *et al.*, S1P-dependent interorgan trafficking of group 2 innate lymphoid cells supports host defense. *Science* **359**, 114–119 (2018). doi: [10.1126/science.aam5809](https://doi.org/10.1126/science.aam5809); pmid: [29302015](https://pubmed.ncbi.nlm.nih.gov/29302015/)
 95. J. Michaud, D. S. Im, T. Hla, Inhibitory role of sphingosine 1-phosphate receptor 2 in macrophage recruitment during inflammation. *J. Immunol.* **184**, 1475–1483 (2010). doi: [10.1049/jimmunol.0901586](https://doi.org/10.1049/jimmunol.0901586); pmid: [20042570](https://pubmed.ncbi.nlm.nih.gov/20042570/)
 96. L. van der Weyden *et al.*, Genome-wide in vivo screen identifies novel host regulators of metastatic colonization. *Nature* **541**, 233–236 (2017). doi: [10.1038/nature20792](https://doi.org/10.1038/nature20792); pmid: [28052056](https://pubmed.ncbi.nlm.nih.gov/28052056/)
 97. A. Mendoza *et al.*, Lymphatic endothelial S1P promotes mitochondrial function and survival in naive T cells. *Nature* **546**, 158–161 (2017). doi: [10.1038/nature22352](https://doi.org/10.1038/nature22352); pmid: [28538737](https://pubmed.ncbi.nlm.nih.gov/28538737/)
 98. S. A. Jaigirdar *et al.*, Sphingosine-1-phosphate promotes the persistence of activated CD4 T cells in inflamed sites. *Front. Immunol.* **8**, 1627 (2017). doi: [10.3389/fimmu.2017.01627](https://doi.org/10.3389/fimmu.2017.01627); pmid: [29225602](https://pubmed.ncbi.nlm.nih.gov/29225602/)
 99. C. S. Garriss *et al.*, Defective sphingosine 1-phosphate receptor 1 (S1P1) phosphorylation exacerbates TH17-mediated autoimmune neuroinflammation. *Nat. Immunol.* **14**, 1166–1172 (2013). doi: [10.1038/ni.2730](https://doi.org/10.1038/ni.2730); pmid: [24076635](https://pubmed.ncbi.nlm.nih.gov/24076635/)
 100. N. Claes *et al.*, Compositional changes of B and T cell subtypes during fingolimod treatment in multiple sclerosis patients: A 12-month follow-up study. *PLOS ONE* **9**, e111115 (2014). doi: [10.1371/journal.pone.0111115](https://doi.org/10.1371/journal.pone.0111115); pmid: [25360562](https://pubmed.ncbi.nlm.nih.gov/25360562/)
 101. A. Eken *et al.*, S1P₁ deletion differentially affects TH17 and regulatory T cells. *Sci. Rep.* **7**, 12905 (2017). doi: [10.1038/s41598-017-13376-2](https://doi.org/10.1038/s41598-017-13376-2); pmid: [29018225](https://pubmed.ncbi.nlm.nih.gov/29018225/)
 102. E. Degagné *et al.*, Sphingosine-1-phosphate lyase downregulation promotes colon carcinogenesis through STAT3-activated microRNAs. *J. Clin. Invest.* **124**, 5368–5384 (2014). doi: [10.1172/JCI74188](https://doi.org/10.1172/JCI74188); pmid: [25347472](https://pubmed.ncbi.nlm.nih.gov/25347472/)
 103. J. Liang *et al.*, Sphingosine-1-phosphate links persistent STAT3 activation, chronic intestinal inflammation, and development of colitis-associated cancer. *Cancer Cell* **23**, 107–120 (2013). doi: [10.1016/j.ccr.2012.11.013](https://doi.org/10.1016/j.ccr.2012.11.013); pmid: [23273921](https://pubmed.ncbi.nlm.nih.gov/23273921/)
 104. Z. Zhao, A. R. Nelson, C. Bethsholtz, B. V. Zlokovic, Establishment and dysfunction of the blood-brain barrier. *Cell* **163**, 1064–1078 (2015). doi: [10.1016/j.cell.2015.10.067](https://doi.org/10.1016/j.cell.2015.10.067); pmid: [26590417](https://pubmed.ncbi.nlm.nih.gov/26590417/)
 105. K. Yanagida *et al.*, Size-selective opening of the blood-brain barrier by targeting endothelial sphingosine 1-phosphate receptor 1. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4531–4536 (2017). doi: [10.1073/pnas.1618659114](https://doi.org/10.1073/pnas.1618659114); pmid: [28396408](https://pubmed.ncbi.nlm.nih.gov/28396408/)
 106. R. E. Cannon, J. C. Peart, B. T. Hawkins, C. R. Campos, D. S. Miller, Targeting blood-brain barrier sphingolipid signaling reduces basal P-glycoprotein activity and improves drug delivery to the brain. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15930–15935 (2012). doi: [10.1073/pnas.1203534109](https://doi.org/10.1073/pnas.1203534109); pmid: [22949658](https://pubmed.ncbi.nlm.nih.gov/22949658/)
 107. W. Pfeilschifter *et al.*, Activation of sphingosine kinase 2 is an endogenous protective mechanism in cerebral ischemia. *Biochem. Biophys. Res. Commun.* **413**, 212–217 (2011). doi: [10.1016/j.bbrc.2011.08.070](https://doi.org/10.1016/j.bbrc.2011.08.070); pmid: [21872577](https://pubmed.ncbi.nlm.nih.gov/21872577/)
 108. G. S. Kim *et al.*, Critical role of sphingosine-1-phosphate receptor-2 in the disruption of cerebrovascular integrity in experimental stroke. *Nat. Commun.* **6**, 7893 (2015). doi: [10.1038/ncomms8893](https://doi.org/10.1038/ncomms8893); pmid: [26243335](https://pubmed.ncbi.nlm.nih.gov/26243335/)
 109. W. B. Rolland *et al.*, Fingolimod reduces cerebral lymphocyte infiltration in experimental models of rodent intracerebral hemorrhage. *Exp. Neurol.* **241**, 45–55 (2013). doi: [10.1016/j.jexpneurol.2012.12.009](https://doi.org/10.1016/j.jexpneurol.2012.12.009); pmid: [23261767](https://pubmed.ncbi.nlm.nih.gov/23261767/)
 110. Y. Fu *et al.*, Impact of an immune modulator fingolimod on acute ischemic stroke. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18315–18320 (2014). doi: [10.1073/pnas.1416166111](https://doi.org/10.1073/pnas.1416166111); pmid: [25489101](https://pubmed.ncbi.nlm.nih.gov/25489101/)
 111. L. Cruz-Orrego *et al.*, Enhanced sphingosine-1-phosphate receptor 2 expression underlies female CNS autoimmunity susceptibility. *J. Clin. Invest.* **124**, 2571–2584 (2014). doi: [10.1172/JCI73408](https://doi.org/10.1172/JCI73408); pmid: [24812668](https://pubmed.ncbi.nlm.nih.gov/24812668/)
 112. S. S. Dusaban, J. Chun, H. Rosen, N. H. Purcell, J. H. Brown, Sphingosine 1-phosphate receptor 3 and RhoA signaling mediate inflammatory gene expression in astrocytes. *J. Neuroinflammation* **14**, 111 (2017). doi: [10.1186/s12974-017-0882-x](https://doi.org/10.1186/s12974-017-0882-x); pmid: [28577576](https://pubmed.ncbi.nlm.nih.gov/28577576/)
 113. J. W. Choi *et al.*, FTY720 (fingolimod) efficacy in an animal model of multiple sclerosis requires astrocyte sphingosine 1-phosphate receptor 1 (S1P1) modulation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 751–756 (2011). doi: [10.1073/pnas.1014154108](https://doi.org/10.1073/pnas.1014154108); pmid: [21177428](https://pubmed.ncbi.nlm.nih.gov/21177428/)
 114. I. Fischer *et al.*, Sphingosine kinase 1 and sphingosine 1-phosphate receptor 3 are functionally upregulated on astrocytes under pro-inflammatory conditions. *PLOS ONE* **6**, e23905 (2011). doi: [10.1371/journal.pone.0023905](https://doi.org/10.1371/journal.pone.0023905); pmid: [21887342](https://pubmed.ncbi.nlm.nih.gov/21887342/)
 115. S. A. Liddelow *et al.*, Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017). doi: [10.1038/nature21029](https://doi.org/10.1038/nature21029); pmid: [28099414](https://pubmed.ncbi.nlm.nih.gov/28099414/)
 116. J. Ceccom *et al.*, Reduced sphingosine kinase-1 and enhanced sphingosine 1-phosphate lyase expression demonstrate deregulated sphingosine 1-phosphate signaling in Alzheimer's disease. *Acta Neuropathol. Commun.* **2**, 12 (2014). doi: [10.1186/2051-5960-2-12](https://doi.org/10.1186/2051-5960-2-12); pmid: [24468113](https://pubmed.ncbi.nlm.nih.gov/24468113/)
 117. T. A. Couttas *et al.*, Loss of the neuroprotective factor Sphingosine 1-phosphate early in Alzheimer's disease pathogenesis. *Acta Neuropathol. Commun.* **2**, 9 (2014). doi: [10.1186/2051-5960-2-9](https://doi.org/10.1186/2051-5960-2-9); pmid: [24456642](https://pubmed.ncbi.nlm.nih.gov/24456642/)
 118. Y. V. Taguchi *et al.*, Glucosylsphingosine promotes α-synuclein pathology in mutant gba-associated Parkinson's disease. *J. Neurosci.* **37**, 9617–9631 (2017). doi: [10.1523/JNEUROSCI.1525-17.2017](https://doi.org/10.1523/JNEUROSCI.1525-17.2017); pmid: [28847804](https://pubmed.ncbi.nlm.nih.gov/28847804/)

119. A. Di Pardo, V. Maglione, The S1P Axis: New exciting route for treating Huntington's disease. *Trends Pharmacol. Sci.* **39**, 468–480 (2018). doi: [10.1016/j.tips.2018.02.009](https://doi.org/10.1016/j.tips.2018.02.009); pmid: [29559169](https://pubmed.ncbi.nlm.nih.gov/29559169/)
120. L. Zhang *et al.*, Extracellular α -synuclein induces sphingosine 1-phosphate receptor subtype 1 uncoupled from inhibitory G-protein leaving β -arrestin signal intact. *Sci. Rep.* **7**, 44248 (2017). doi: [10.1038/srep44248](https://doi.org/10.1038/srep44248); pmid: [28300069](https://pubmed.ncbi.nlm.nih.gov/28300069/)
121. N. Aytan *et al.*, Fingolimod modulates multiple neuroinflammatory markers in a mouse model of Alzheimer's disease. *Sci. Rep.* **6**, 24939 (2016). doi: [10.1038/srep24939](https://doi.org/10.1038/srep24939); pmid: [27117087](https://pubmed.ncbi.nlm.nih.gov/27117087/)
122. S. W. Park, M. Kim, K. M. Brown, V. D. D'Agati, H. T. Lee, Inhibition of sphingosine 1-phosphate receptor 2 protects against renal ischemia-reperfusion injury. *J. Am. Soc. Nephrol.* **23**, 266–280 (2012). doi: [10.1681/ASN.2011050503](https://doi.org/10.1681/ASN.2011050503); pmid: [22095950](https://pubmed.ncbi.nlm.nih.gov/22095950/)
123. S. Thangada *et al.*, Treatment with the immunomodulator FTY720 (fingolimod) significantly reduces renal inflammation in murine unilateral ureteral obstruction. *J. Urol.* **191** (Suppl), 1508–1516 (2014). doi: [10.1016/j.juro.2013.10.072](https://doi.org/10.1016/j.juro.2013.10.072); pmid: [24679864](https://pubmed.ncbi.nlm.nih.gov/24679864/)
124. S. Schwalm *et al.*, Sphingosine kinase-2 deficiency ameliorates kidney fibrosis by up-regulating Smad7 in a mouse model of unilateral ureteral obstruction. *Am. J. Pathol.* **187**, 2413–2429 (2017). doi: [10.1016/j.ajpath.2017.06.017](https://doi.org/10.1016/j.ajpath.2017.06.017); pmid: [28807595](https://pubmed.ncbi.nlm.nih.gov/28807595/)
125. M. Ghosh *et al.*, Cell-intrinsic sphingosine kinase 2 promotes macrophage polarization and renal inflammation in response to unilateral ureteral obstruction. *PLOS ONE* **13**, e0194053 (2018). doi: [10.1371/journal.pone.0194053](https://doi.org/10.1371/journal.pone.0194053); pmid: [29518138](https://pubmed.ncbi.nlm.nih.gov/29518138/)
126. A. Bajwa *et al.*, Sphingosine kinase 2 deficiency attenuates kidney fibrosis via IFN- γ . *J. Am. Soc. Nephrol.* **28**, 1145–1161 (2017). doi: [10.1681/ASN.2016030306](https://doi.org/10.1681/ASN.2016030306); pmid: [27799486](https://pubmed.ncbi.nlm.nih.gov/27799486/)
127. M. Sato *et al.*, Sphingosine kinase-1, S1P transporter spinster homolog 2 and S1P2 mRNA expressions are increased in liver with advanced fibrosis in human. *Sci. Rep.* **6**, 32119 (2016). doi: [10.1038/srep32119](https://doi.org/10.1038/srep32119); pmid: [27562371](https://pubmed.ncbi.nlm.nih.gov/27562371/)
128. L. Yang *et al.*, Sphingosine kinase/sphingosine 1-phosphate (S1P)/S1P receptor axis is involved in liver fibrosis-associated angiogenesis. *J. Hepatol.* **59**, 114–123 (2013). doi: [10.1016/j.jhep.2013.02.021](https://doi.org/10.1016/j.jhep.2013.02.021); pmid: [23466305](https://pubmed.ncbi.nlm.nih.gov/23466305/)
129. H. Ikeda *et al.*, Sphingosine 1-phosphate regulates regeneration and fibrosis after liver injury via sphingosine 1-phosphate receptor 2. *J. Lipid Res.* **50**, 556–564 (2009). doi: [10.1194/jlr.M800496-JLR200](https://doi.org/10.1194/jlr.M800496-JLR200); pmid: [18955732](https://pubmed.ncbi.nlm.nih.gov/18955732/)
130. A. Stelling *et al.*, The tumor suppressive TGF- β /SMAD1/S1P2 signaling axis is recurrently inactivated in diffuse large B-cell lymphoma. *Blood* **131**, 2235–2246 (2018). doi: [10.1182/blood-2017-10-810630](https://doi.org/10.1182/blood-2017-10-810630); pmid: [29615404](https://pubmed.ncbi.nlm.nih.gov/29615404/)
131. L. Yang *et al.*, Sphingosine 1-phosphate receptor 2 and 3 mediate bone marrow-derived monocyte/macrophage motility in cholestatic liver injury in mice. *Sci. Rep.* **5**, 13423 (2015). doi: [10.1038/srep13423](https://doi.org/10.1038/srep13423); pmid: [26324256](https://pubmed.ncbi.nlm.nih.gov/26324256/)
132. J. Milara *et al.*, Sphingosine-1-phosphate is increased in patients with idiopathic pulmonary fibrosis and mediates epithelial to mesenchymal transition. *Thorax* **67**, 147–156 (2012). doi: [10.1136/thoraxjnl-2011-200026](https://doi.org/10.1136/thoraxjnl-2011-200026); pmid: [22106015](https://pubmed.ncbi.nlm.nih.gov/22106015/)
133. K. Sobel *et al.*, Sphingosine 1-phosphate (S1P) receptor agonists mediate pro-fibrotic responses in normal human lung fibroblasts via S1P2 and S1P3 receptors and Smad-independent signaling. *J. Biol. Chem.* **288**, 14839–14851 (2013). doi: [10.1074/jbc.M112.426726](https://doi.org/10.1074/jbc.M112.426726); pmid: [23589284](https://pubmed.ncbi.nlm.nih.gov/23589284/)
134. J. Zhao *et al.*, Sphingosine-1-phosphate receptor-2 facilitates pulmonary fibrosis through potentiating IL-13 pathway in macrophages. *PLOS ONE* **13**, e0197604 (2018). doi: [10.1371/journal.pone.0197604](https://doi.org/10.1371/journal.pone.0197604); pmid: [29782549](https://pubmed.ncbi.nlm.nih.gov/29782549/)
135. K. Murakami *et al.*, Knock out of S1P3 receptor signaling attenuates inflammation and fibrosis in bleomycin-induced lung injury mice model. *PLOS ONE* **9**, e106792 (2014). doi: [10.1371/journal.pone.0106792](https://doi.org/10.1371/journal.pone.0106792); pmid: [25198418](https://pubmed.ncbi.nlm.nih.gov/25198418/)
136. A. J. Dyckman, Modulators of Sphingosine-1-phosphate pathway biology: Recent advances of sphingosine-1-phosphate receptor 1 (S1P₁) agonists and future perspectives. *J. Med. Chem.* **60**, 5267–5289 (2017). doi: [10.1021/acs.jmedchem.6b01575](https://doi.org/10.1021/acs.jmedchem.6b01575); pmid: [28291340](https://pubmed.ncbi.nlm.nih.gov/28291340/)
137. L. Kappos *et al.*, Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): A double-blind, randomised, phase 3 study. *Lancet* **391**, 1263–1273 (2018). doi: [10.1016/S0140-6736\(18\)30475-6](https://doi.org/10.1016/S0140-6736(18)30475-6); pmid: [29576505](https://pubmed.ncbi.nlm.nih.gov/29576505/)
138. C. Jaillard *et al.*, Edg8/S1P5: An oligodendroglial receptor with dual function on process retraction and cell survival. *J. Neurosci.* **25**, 1459–1469 (2005). doi: [10.1523/JNEUROSCI.4645-04.2005](https://doi.org/10.1523/JNEUROSCI.4645-04.2005); pmid: [15703400](https://pubmed.ncbi.nlm.nih.gov/15703400/)
139. Z. Rankovic, T. F. Brust, L. M. Bohn, Biased agonism: An emerging paradigm in GPCR drug discovery. *Bioorg. Med. Chem. Lett.* **26**, 241–250 (2016). doi: [10.1016/j.bmcl.2015.12.024](https://doi.org/10.1016/j.bmcl.2015.12.024); pmid: [26707396](https://pubmed.ncbi.nlm.nih.gov/26707396/)
140. E. R. Viscusi *et al.*, A randomized, phase 2 study investigating TRV130, a biased ligand of the μ -opioid receptor, for the intravenous treatment of acute pain. *Pain* **157**, 264–272 (2016). doi: [10.1097/j.pain.0000000000000363](https://doi.org/10.1097/j.pain.0000000000000363); pmid: [26683109](https://pubmed.ncbi.nlm.nih.gov/26683109/)
141. A. Skoura *et al.*, Sphingosine-1-phosphate receptor-2 function in myeloid cells regulates vascular inflammation and atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **31**, 81–85 (2011). doi: [10.1161/ATVBAHA.110.213496](https://doi.org/10.1161/ATVBAHA.110.213496); pmid: [20947824](https://pubmed.ncbi.nlm.nih.gov/20947824/)
142. A. J. MacLennan *et al.*, An essential role for the H218/AGR16/Edg-5/LP(B2) sphingosine 1-phosphate receptor in neuronal excitability. *Eur. J. Neurosci.* **14**, 203–209 (2001). doi: [10.1046/j.0953-816x.2001.01634.x](https://doi.org/10.1046/j.0953-816x.2001.01634.x); pmid: [11553273](https://pubmed.ncbi.nlm.nih.gov/11553273/)
143. A. Billich *et al.*, Partial deficiency of sphingosine-1-phosphate lyase confers protection in experimental autoimmune encephalomyelitis. *PLOS ONE* **8**, e59630 (2013). doi: [10.1371/journal.pone.0059630](https://doi.org/10.1371/journal.pone.0059630); pmid: [23544080](https://pubmed.ncbi.nlm.nih.gov/23544080/)
144. S. Pyne, D. R. Adams, N. J. Pyne, Sphingosine kinases as druggable targets, in *Handbook of Experimental Pharmacology* (Springer, 2018).
145. M. E. Schnute *et al.*, Modulation of cellular S1P levels with a novel, potent and specific inhibitor of sphingosine kinase-1. *Biochem. J.* **444**, 79–88 (2012). doi: [10.1042/BJ20111929](https://doi.org/10.1042/BJ20111929); pmid: [22397330](https://pubmed.ncbi.nlm.nih.gov/22397330/)
146. Y. Kharel *et al.*, Sphingosine kinase 2 inhibition and blood sphingosine 1-phosphate levels. *J. Pharmacol. Exp. Ther.* **355**, 23–31 (2015). doi: [10.1124/jpet.115.225862](https://doi.org/10.1124/jpet.115.225862); pmid: [26243740](https://pubmed.ncbi.nlm.nih.gov/26243740/)

ACKNOWLEDGMENTS

Funding: This work is supported by NIH grant R35 HL135821 (to T.H.), Fondation Leducq transatlantic network grant (SphingoNet) (to T.H.), and a postdoctoral fellowship from the American Heart Association (to A.C.). **Author contributions:** A.C. and T.H. both wrote and edited the manuscript. A.C. prepared Figs. 1 to 5 and compiled Table 1. **Competing interests:** T.H. received a research grant from ONO Pharmaceutical; is an inventor in the patents and patent applications related to ApoM-Fc, ApoM⁺-HDL, and S1P receptor modulators; and has consulted for the following commercial entities: Abbott, Pfizer, Sandoz, Astellas, Novartis, Trevana, and Arena Pharmaceuticals.

10.1126/science.aar5551

RESEARCH ARTICLE SUMMARY

HUMAN EVOLUTION

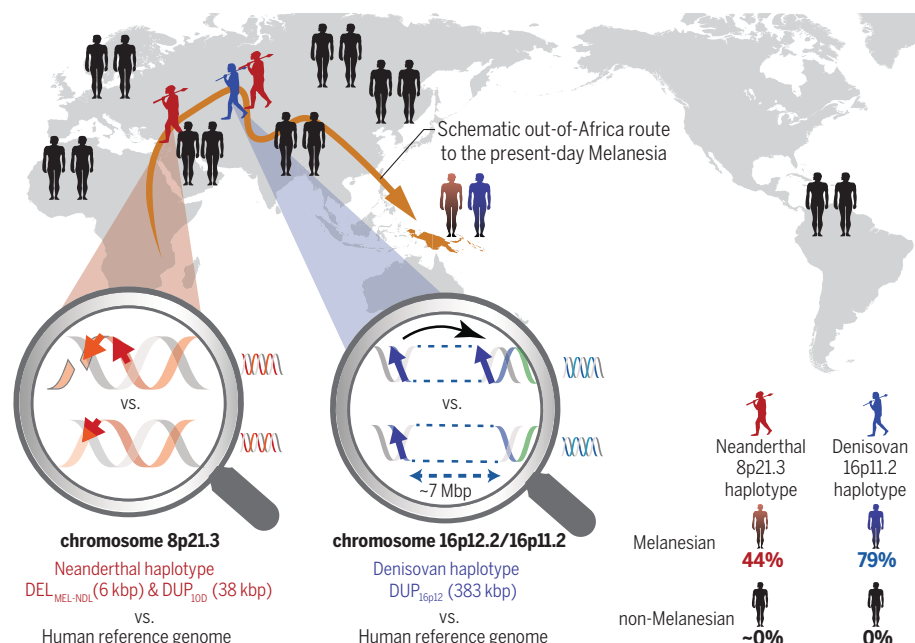
Adaptive archaic introgression of copy number variants and the discovery of previously unknown human genes

PingHsun Hsieh, Mitchell R. Vollger, Vy Dang, David Porubsky, Carl Baker, Stuart Cantsilieris, Kendra Hoekzema, Alexandra P. Lewis, Katherine M. Munson, Melanie Sorensen, Zev N. Kronenberg, Shwetha Murali, Bradley J. Nelson, Giorgia Chiatante, Flavia Angela Maria Maggiolini, Hélène Blanché, Jason G. Underwood, Francesca Antonacci, Jean-François Deleuze, Evan E. Eichler*

INTRODUCTION: Characterizing genetic variants underlying local adaptations in human populations is one of the central goals of evolutionary research. Most studies have focused on adaptive single-nucleotide variants that either arose as new beneficial mutations or were introduced after interbreeding with our now-extinct relatives, including Neanderthals and Denisovans. The adaptive role of copy number variants (CNVs), another well-known form of genomic variation generated through deletions or duplications that affect more base pairs in the genome, is less well understood, despite evidence that such mutations are subject to stronger selective pressures.

RATIONALE: This study focuses on the discovery of introgressed and adaptive CNVs that

have become enriched in specific human populations. We combine whole-genome CNV calling and population genetic inference methods to discover CNVs and then assess signals of selection after controlling for demographic history. We examine 266 publicly available modern human genomes from the Simons Genome Diversity Project and genomes of three ancient hominins—a Denisovan, a Neanderthal from the Altai Mountains in Siberia, and a Neanderthal from Croatia. We apply long-read sequencing methods to sequence-resolve complex CNVs of interest specifically in the Melanesians—an Oceanian population distributed from Papua New Guinea to as far east as the islands of Fiji and known to harbor some of the greatest amounts of Neanderthal and Denisovan ancestry.



Large adaptive-introgressed CNVs at chromosomes 8p21.3 and 16p11.2 in Melanesians. The magnifying glasses highlight structural differences between the archaic (top) and reference (bottom) genomes. Neanderthal (red) and Denisovan (blue) haplotypes encompassing large CNVs occur at high frequencies in Melanesians (44 and 79%, respectively) but are absent (black) in all non-Melanesians. These CNVs create positively selected genes (*TNFRSF10D1*, *TNFRSF10D2*, and *NPIP16*) that are absent from the reference genome.

RESULTS: Consistent with the hypothesis of archaic introgression outside Africa, we find a significant excess of CNV sharing between modern non-African populations and archaic hominins ($P = 0.039$). Among Melanesians, we observe an enrichment of CNVs with potential signals of positive selection ($n = 37$ CNVs), of which 19 CNVs likely introgressed from archaic hominins. We show that Melanesian-stratified CNVs are significantly associated with signals of positive selection ($P = 0.0323$). Many map near or within genes associated with metabolism (e.g., *ACOT1* and *ACOT2*), development and cell cycle or signaling (e.g., *TNFRSF10D* and *CDK11A* and *CDK11B*), or immune response (e.g., *IFNLRI*). We characterize two of the largest and most complex CNVs on chromosomes 16p11.2 and 8p21.3 that introgressed from Denisovans and Neanderthals, respectively, and are absent from most other human populations. At chromosome 16p11.2, we sequence-resolve a large duplication of >383 thousand base pairs (kbp) that originated from Denisovans and introgressed into the ancestral Melanesian population 60,000 to 170,000 years ago. This large duplication occurs at high frequency (>79%) in diverse Melanesian groups, shows signatures of positive selection, and maps adjacent to *Homo sapiens*-specific duplications that predispose to rearrangements associated with autism. On chromosome 8p21.3, we identify a Melanesian haplotype that carries two CNVs, a ~6-kbp deletion, and a ~38-kbp duplication, with a Neanderthal origin and that introgressed into non-Africans 40,000 to 120,000 years ago. This CNV haplotype occurs at high frequency (44%) and shows signals consistent with a partial selective sweep in Melanesians. Using long-read sequencing genomic and transcriptomic data, we reconstruct the structure and complex evolutionary history for these two CNVs and discover previously undescribed duplicated genes (*TNFRSF10D1*, *TNFRSF10D2*, and *NPIP16*) that show an excess of amino acid replacements consistent with the action of positive selection.

CONCLUSION: Our results suggest that large CNVs originating in archaic hominins and introgressed into modern humans have played an important role in local population adaptation and represent an insufficiently studied source of large-scale genetic variation that is absent from current reference genomes. ■

The list of author affiliations is available in the full article online.

*Corresponding author. Email: eee@gs.washington.edu

Cite this article as P. Hsieh et al., *Science* 366, eaax2083 (2019). DOI: 10.1126/science.aax2083

RESEARCH ARTICLE

HUMAN EVOLUTION

Adaptive archaic introgression of copy number variants and the discovery of previously unknown human genes

PingHsun Hsieh¹, Mitchell R. Vollger¹, Vy Dang¹, David Porubsky¹, Carl Baker¹, Stuart Cantsilieris^{1*}, Kendra Hoekzema¹, Alexandra P. Lewis¹, Katherine M. Munson¹, Melanie Sorensen¹, Zev N. Kronenberg^{1†}, Shwetha Murali^{1,2}, Bradley J. Nelson^{1‡}, Giorgia Chiatante³, Flavia Angela Maria Maggolini³, Hélène Blanché⁴, Jason G. Underwood^{1,5}, Francesca Antonacci³, Jean-François Deleuze^{4,5}, Evan E. Eichler^{1,2§}

Copy number variants (CNVs) are subject to stronger selective pressure than single-nucleotide variants, but their roles in archaic introgression and adaptation have not been systematically investigated. We show that stratified CNVs are significantly associated with signatures of positive selection in Melanians and provide evidence for adaptive introgression of large CNVs at chromosomes 16p11.2 and 8p21.3 from Denisovans and Neanderthals, respectively. Using long-read sequence data, we reconstruct the structure and complex evolutionary history of these polymorphisms and show that both encode positively selected genes absent from most human populations. Our results collectively suggest that large CNVs originating in archaic hominins and introgressed into modern humans have played an important role in local population adaptation and represent an insufficiently studied source of large-scale genetic variation.

Upon migration out of their ancestral homeland, our ancestors likely experienced pressure to adapt to new environments. Genomic surveys with single-nucleotide variants (SNVs) have provided evidence for local genetic adaptations (1–4) and the acquisition of variants introgressed from other hominins and subsequently under positive selection in modern humans (5–7). Copy number variants (CNVs) are a form of structural variation, including deletions, duplications, and multi-allelic CNVs, that are >50 base pairs (bp) in length and affect at least five times as many base pairs as SNVs per diploid human genome (8–10). When compared with the relative abundance of SNVs, structural variants are three times as likely to be associated with a genome-wide association signal and 50 times as likely to be associated with expressed quantitative trait loci (eQTL) (11). Although larger CNVs are generally deleterious and associated with disease, examples of adaptive CNVs in humans have been documented (12–15). However, relatively little is known about the ex-

tent to which CNVs contribute to the genetic basis of local adaptation and, more importantly, whether CNVs introgressed from other hominins may have been targets of adaptive selection.

In this work, we systematically searched for genome-wide evidence for selective and archaic introgressed CNVs among Melanesian genomes. Melanesia, a subregion of Oceania, is characterized by island populations distributed across the Bismarck Archipelago and Bougainville Island, just to the east of New Guinea. Melanians have likely developed adaptations to diet (16), infectious diseases (17), and body size (18) as a result of the tropical island environment. In addition, Melanians have been relatively isolated over most of their history (>50,000 years), with major influences from populations to the west occurring only during the past 3000 years (19, 20). Melanians also carry some of the greatest amounts of ancestry from both known archaic hominins, Denisovans (3 to 5%) (21–23) and Neanderthals (1 to 3%) (24), allowing for insights into the role of introgression in human evolution.

An excess of archaic hominin CNVs in non-African populations

We began by constructing a database of 5135 archaic CNVs (table S1) from previous sequence read-depth analyses of genomes of three ancient hominins—a Denisovan (25), a Neanderthal (26) from the Altai Mountains in Siberia, and a Neanderthal from Croatia (24). We compared these archaic genomes to a diversity panel of 17 high-coverage modern human genomes from the Simons Genome Diversity

Project (SGDP) (24). We genotyped these CNVs in 249 genomes from the SGDP panel (27) and 72 nonhuman great ape genomes (28) and then parsimoniously classified CNV events as lineage-specific or shared with modern humans and other ape species. We found that 2684 of the 5135 CNV loci in modern humans are copy number polymorphic in the archaic genomes [i.e., at least one of the three archaic genomes has a copy number estimate not equal to 2]; of these, 142 CNVs are likely derived relative to the 72 nonhuman primate genomes, which are fixed at a diploid copy number (CN2) for these loci. We defined a CNV as hominin specific if it is fixed in CN2 among all nonhuman primate samples and is variable in copy number in at least one hominin sample. Among the 402 hominin-specific CNVs (table S2), 13% (51 of 402) of the CNVs are specifically shared between the three Eurasian archaic hominins and non-African samples. The amount of shared CNVs between the archaic and non-African samples is unlikely under a null expectation ($P = 0.039$, 100,000 permutation simulations; fig. S1). Despite the small sample size of three archaic genomes, our result is consistent with recent evidence for archaic admixture found in Eurasia and suggests that some of the shared CNVs might result from archaic introgression.

Discovery of single-nucleotide and population-stratified CNVs

We performed a more comprehensive computational analysis to discover and assess patterns of single-nucleotide and structural variation in the SGDP populations (29). We identified population-stratified CNVs and then tested for selection and introgression by using SNVs from the diploid sequence flanking population-stratified CNVs (29). Using five CNV callers, we discovered 368,256 autosomal CNVs, of which 93.5% were identified by a single CNV caller (fig. S2 and table S3). We focused on a conserved call set of 19,211 CNVs constructed by applying a variety of quality control filters (fig. S3) (29). We used orthogonal single-nucleotide polymorphism microarray data to determine an overall validation rate of 83.5% (68.2% for duplications and ~100% for deletions) (table S4) (29).

To identify population-stratified CNVs, we applied three statistics quantifying the differences in copy number between a focal population and all the other SGDP samples: V_{ST} , Mann-Whitney U (MWU), and D_{median} (29). After removing admixed individuals (fig. S4), we grouped the remaining SGDP samples ($n = 249$ individuals) into eight focal populations: sub-Saharan Africans ($n = 33$), Native Americans ($n = 20$), East Asians ($n = 47$), Europeans ($n = 51$), Melanians ($n = 16$), Middle Easterners ($n = 22$), South Asians ($n = 38$), and Siberians ($n = 22$) (29). We found little to no

¹Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA. ²Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.

³Dipartimento di Biologia, Università degli Studi di Bari "Aldo Moro," Bari, Italy. ⁴Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain, Paris, France. ⁵Pacific Biosciences (PacBio) of California, Inc., Menlo Park, CA, USA.

*Present address: Centre for Eye Research Australia, Department of Surgery (Ophthalmology), University of Melbourne, Royal Victorian Eye and Ear Hospital, East Melbourne, VIC, Australia.

†Present address: Pacific Biosciences (PacBio) of California, Inc., Menlo Park, CA, USA. ‡Present address: Phase Genomics, Inc., Seattle, WA, USA.

§Corresponding author. Email: eee@gs.washington.edu

correlation between the focal population sample size and the number of stratified CNVs identified by V_{ST} ($r = -0.16$, $P = 0.699$) and MWU test ($r = -0.19$, $P = 0.638$) compared with those obtained by the D_{median} test ($r = -0.73$, $P = 0.039$) (fig. S5). In addition, there is a strong correlation between the Bonferroni P values of the MWU and V_{ST} tests ($r = -0.65$, $P < 2 \times 10^{-16}$); however, 233 CNVs have $V_{ST} > 0.1$ but with MWU Bonferroni's $P > 0.05$, indicating that the two tests are complementary. To conservatively identify stratified CNVs in a focal population, we used the following criteria: (i) $V_{ST} > 0.1$, (ii) Bonferroni P value of the MWU test < 0.05 , and (iii) $D_{median} > 0.5$ (fig. S6). Although the number varied by population (table S5), we observed more stratified CNVs per population than expected by chance ($P < 0.0105$, 10,000 nonparametric permutation simulations) (29), suggesting that the enrichment of stratified CNVs is unlikely to be the result of sampling errors. Because Melanesians share the most introgressed alleles with archaic lineages (an estimated 2 to 4% from both Denisovan and Neanderthal), we focused on this population for a more detailed investigation of introgressed and selected CNVs.

Tests for selective and archaic introgressed CNVs in Melanesians

We considered two possible origins for the population-stratified CNVs: The events could have arisen de novo or, alternatively, could have introgressed from other hominins and then subsequently risen to high frequency by the action of natural selection or demographic processes, including drift. To distinguish between these hypotheses, we applied a variety of population genetic statistics that use both allele frequency and linkage disequilibrium information of SNVs flanking the CNVs for additional evidence of archaic hominin introgression and/or signatures of positive selection (29). Potential biases on the test statistics owing to population history were controlled through coalescent simulations on the basis of the inferred best-fit demographic models for Melanesians. We define the significance of each test as the fraction of simulations with test statistic values greater than or equal to the observed value in the real data (29). A test for a window is significant if its P value is < 0.05 . Among the models we tested (table S6), our best-fit model (fig. S7; log-likelihood = $-103,386$) estimates that the ancestors of Africans and non-Africans diverged ~74,000 years ago [95% confidence interval (CI): 73,321 to 75,199 years ago], followed by Melanesian–East Asian divergence ~52,000 years ago (95% CI: 51,450 to 52,591 years ago), consistent with recent reports (19). Our demographic model confirms moderately high gene flow from East Asians to Melanesians [95% CI for $N_{AMH} * m_{MEL-EA} = 1.109$ to $1.128 > 1$ (N_{AMH} , effective population

size of anatomically modern humans; m_{MEL-EA} , migration rate from East Asia to Melanesia); table S7]. This finding is notable because there is evidence for East Asian gene flow into Melanesians, especially in the lowland groups, which constitute approximately one-third of our Melanesian sample set (19, 20).

Our coalescent simulations (fig. S8 and table S8) (29) generally recapitulated the empirical variation pattern of SNVs (fig. S9A) and demonstrated the utility of our approach to identify non-neutrally evolved loci (fig. S9, B and C). We used these parametric coalescent simulations to estimate significance of selection and introgression at individual loci. Among Melanesians, we identified signatures of positive selection at 37 distinct CNV loci [$P < 0.05$, population branch statistic (PBS)] (2) (Fig. 1) and signals of introgression [f_D test (30) at 24 and 28 CNV loci using Neanderthal and Denisovan genomes as archaic references, respectively]. Notably, more than 51% of the selective CNV candidates (19 of 37 CNVs) also have introgression signals at the flanking diploid sequences (Table 1 and tables S9 and S10)—an observation that is highly unlikely ($P = 0.0004$, 10,000 nonparametric permutations). However, we hypothesize that this strong correlation is likely driven by the association between population-stratified CNVs and selection signals at their flanking sequences, because our analysis shows that Melanesian-stratified CNVs are significantly associated with selective signals ($P = 0.0323$, 10,000 nonparametric permutations) but not with archaic introgression signals ($P = 0.2282$, 10,000 nonparametric permutations) (fig. S10).

Among the adaptive introgressed CNV candidates, many map near or within genes associated with metabolism (e.g., *ACOT1* and *ACOT2*), development and cell cycle or cell signaling (e.g., *TNFRSF10D*, *CDK11A*, and *CDK11B*), or immune response (e.g., *IFNLRI*); however, the patterns are complex. For example, whereas the *ACOT1* locus harbors a deletion polymorphism common in East Asians (31, 32), we find a ~30-kbp duplication of *ACOT1* in only 12 SGDP samples, including three Melanesians, thus indicating a multi-allelic CNV (figs. S11 and S12). *ACOT1* is involved in regulating cellular balance between free fatty acids and acyl-CoAs, which are critical for energy expenditure and neuronal function (31). Although the *ACOT1* deletion allele is common across populations, we found that >84% of Melanesians retain the *ACOT1* sequences (fig. S12). Furthermore, the major Melanesian haplotype occurs at a much lower frequency in other populations (<11%) and shows extended haplotype homozygosity (EHH), a signature of positive selection (figs. S11, S13, and S14). Although a similar EHH pattern was observed in other non-Africans, most individuals in those populations carry a different allele at the core

(figs. S11 and S14). Notably, the introgression signal in Melanesians at this region is located on the minor haplotype, not the major haplotype where we observe the strongest signature of selection (figs. S13 and S14). The core SNVs (rs4903119 and rs8015976; fig. S11) on the major haplotype are also associated with eQTL differences in *ACOT1*, *ACOT2*, and the adjacent gene, *HEATR4* (GTEx Portal accessed on 24 June 2019; <https://gtexportal.org>), although additional work will be required to fully define these associations and their potential biological consequences.

It is also noteworthy that our analysis identified CNVs and structural variants with strong signals of selection that have been observed in other human populations (table S9) (29). For example, we identify a deletion-linked haplotype at the *APOEC3A* and *APOEC3B* locus that is fixed in Melanesians but relatively rare in other populations (figs. S15 to S19). This deletion allele has been reported at high frequency among Oceanians (32) and is associated with the risk of HIV-1 acquisition (33) as well as breast and ovarian cancers in some Eurasian populations (34). At another immunity-related locus, the alpha-defensin (*DEFA1-T1*) gene family, we find that 87.5% of Melanesians have more *DEFA1-T1* copies than other populations (figs. S20 to S23). They also carry a distinctive haplotype that is rare elsewhere (figs. S23 to S25). This locus is one of the most structurally dynamic regions in the human genome, where recurrent rearrangements have occurred during primate evolution and have been associated with diseases such as immunoglobulin A nephropathy (35, 36). At chromosome 17q21.31, we observe signals of selection and introgression. A large inversion and recurrent duplications in this region have been associated with positive selection and predisposition to disease in Europeans (37, 38). Among Melanesians, we find a nearly fixed haplotype (97%) that is more reminiscent of the ancestral haplotype, which is devoid of both the large inversion and the duplication alleles (figs. S26 to S28). The Melanesian haplotype shows multiple signals of positive selection, including significantly negative Tajima's D (-2.14 , $P = 0.007$), low nucleotide diversity ($\pi < 4.3 \times 10^{-5}$, $P = 0.042$), and elevated EHH across this region (figs. S29 to S32).

Although each of these loci will require more detailed investigation, we focus here on two of the largest and most complex copy number polymorphisms discovered among the Melanesians. Given that the DNA of the Melanesian genomes in the SGDP panel comes from cell lines, which are subject to somatic artifacts, we designed and performed experiments using additional blood-derived Melanesian DNA samples for several CNVs (fig. S33), as well as the two most complex CNV loci reported here, to safeguard against cell line artifacts.

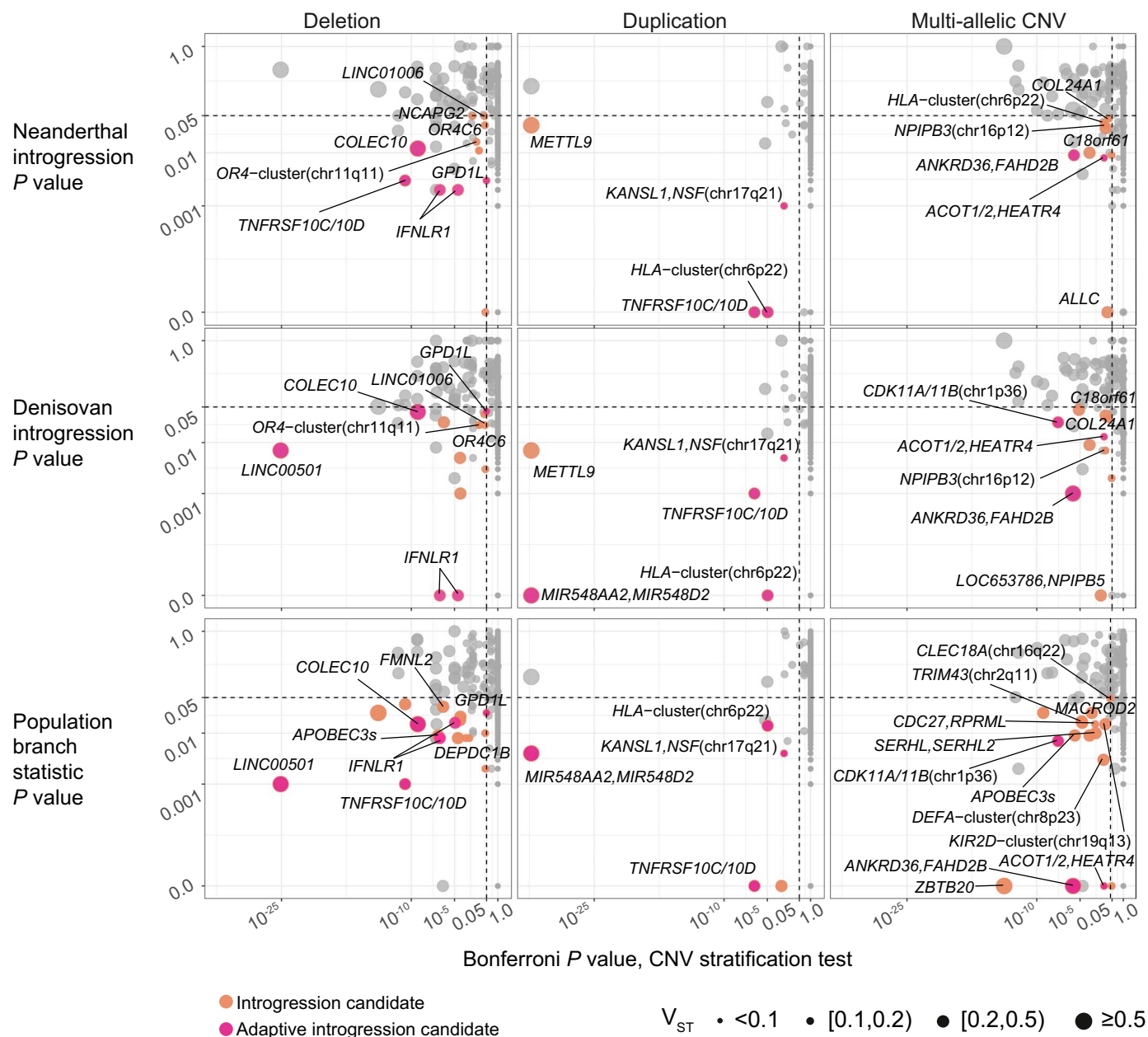


Fig. 1. Candidates for introgressed archaic and selective CNVs. Joint distributions of P values for CNV stratification (x axis, Mann-Whitney U), archaic introgression (f_0 statistic, top two rows), and positive selection (population branch statistic, bottom row) tests. The archaic reference sequences used in the calculations of f_0 are Neanderthals (top row) and Denisovans (middle row), respectively. CNVs that show signatures of both positive selection and archaic introgression (red circles) are distinguished from loci that show signatures of introgression only (orange circles).

Adaptive introgression of the Melanesian–Denisovan–specific duplication at chromosome 16p11.2

The top two Melanesian-stratified CNVs ($P = 2.5 \times 10^{-32}$) are a 5-kbp (chr16:21,596,722–21,601,720) duplication variant and a 73.5-kbp (chr16:22,710,041–22,783,558) duplication variant (fig. S34) originating from chromosome 16p12.2. Both events are largely restricted to Melanesians and the Denisovan archaic genome (Fig. 2A) and are thought to be involved in a single >225-kbp complex duplication (DUP_{16p12})

introgressed from the Denisovan genome (9). This region has been difficult to correctly sequence and assemble, and only recently has the sequence structure of the ancestral locus (>1.1 Mbp) been correctly resolved [KV880768.1, National Center for Biotechnology Information (NCBI) BioProject: PRJNA31257; fig. S35] (39). Unless otherwise stated, we use this patch of sequence to define the ancestral locus.

By using a series of polymerase chain reaction (PCR)–restriction digest assays to genotype an independent set of 242 blood-derived

DNA samples from diverse Melanesians across eight different population groups (Fig. 2B and table S11), we confirmed that this duplication allele is present at high frequency in Melanesians. Although DUP_{16p12} is present at high frequency in all groups [allele frequency (AF) > 0.79], introgression is nearly complete among lowland populations of West and East New Britain, Papua New Guinea (Fig. 2B and table S11). Leveraging paralogous sequence variants that distinguish the duplicated loci, we applied the f_0 statistic and found significant signals for

Table 1. Top candidates for adaptive introgressed CNVs in Melanesians. *P* values for selection and introgression scans are calculated from coalescent simulations. Only loci with selection test *P* values <0.01 are listed. Variant type is determined with respect to the human reference genome. The introgression statistic (*f_D*) is calculated separately for Neanderthals (NDL) and Denisovans (DNS). mCNV, multi-allelic CNV; DUP, bi-allelic duplication; DEL, bi-allelic deletion. *V_{ST}* is a measurement of copy number stratification. See tables S9 and S10 for all candidates.

Locus	Size (bp)	Type	Median CN (Melanesians)	Median CN (all others)	<i>V_{ST}</i>	<i>P</i> value selection (PBS)	<i>P</i> value, introgression (<i>f_D</i>) archaic ref: NDL	<i>P</i> value, introgression (<i>f_D</i>) archaic ref: DNS	Genes within 30 kbp
chr2:97699365-97907043	207,678	mCNV	48.722	39.147	0.487	<0.001*	0.009*	0.001*	ANKRD36, FAHD2B
chr8:22886278-23003689†	117,411	DUP	4.054	3.257	0.272	<0.001*	0.003*	0.055	TNFRSF10D/10C
chr14:73990658-74053946	63,288	mCNV	3.997	2.968	0.149	<0.001*	0.008*	0.013*	ACOT1/2, HEATR4
chr3:177002637-177011375	8738	DEL	0.984	1.905	0.702	0.001*	0.363	0.007*	LINC00501
chr8:22981867-22988907	7040	DEL	1.160	2.163	0.428	0.001*	<0.001*	0.001*	TNFRSF10D/10C
chr16:22710041-22783558††	73,517	DUP	3.935	1.969	0.923	0.004*	0.179	<0.001*	MIR548AA2, MIR548D2
chr17:44170850-45157111	986,261	DUP	2.079	3.311	0.172	0.004*	0.001*	0.005*	KANSL1, NSF(chr17q21)
chr1:1568057-1683771	115,714	mCNV	3.291	4.306	0.298	0.007*	0.053	0.025*	CDK11A/11B(chr1p36)
chr1:24520349-24523673	3324	DEL	0.790	1.898	0.353	0.008*	0.002*	<0.001*	IFNLRI

*Significant test result.

†Newly identified genes: *TNFRSF10D1* and *TNFRSF10D2*.

††Newly identified gene: *NPIPBI6*.

archaic introgression at both loci ($P < 0.033$ and < 0.001 ; Fig. 1, Table 1, and table S10) using Denisovans, but not Neanderthals ($P = 0.178$; fig. S36), as the source of admixture.

Because the elevated PBS and Tajima's D at 16p12.2 (fig. S37) are driven by paralogous sequence variants that are not fully resolved, we sought to further investigate evidence for selection of DUP_{16p12} at its true genomic integration site. We performed a series of fluorescence in situ hybridization (FISH) experiments using probes tiled along the short arm of chromosome 16 in Melanesian cell lines (Fig. 2C, figs. S38 and S39, and tables S12 and S13) (29). Our data localized the derived Melanesian duplication to chromosome 16p11.2 between genomic coordinates of 28.93 and 30.09 Mbp (Fig. 2D). In addition, we generated long-read, high-coverage (75×) whole-genome sequences from a Melanesian individual (HGDP00550; cell line from the HGDP-CEPH panel) who carries the DUP_{16p12} variant. Using the long-read data, we further narrowed the insertion location of this duplication to a 200-kbp interval (29.48 to 29.68 Mbp) adjacent to an *NPIP* (nuclear pore-interacting protein) core duplcon (fig. S40) (29). This is a complex region of >500 kbp of segmental duplication where a *Homo sapiens*-specific duplication emerged ~280,000 years ago and, as a result, predisposes the region to recurrent structural rearrangements associated with autism and developmental delay (15, 40).

To sequence-resolve the DUP_{16p12} copy number polymorphism, we generated a Melanesian large-insert bacterial artificial chromosome (BAC) library (GM10539). From these BACs, we constructed two haplotypes of 222 and 133 kbp, partially confirming the structure of DUP_{16p12} (9) (fig. S41 and S42). To fully assemble the entire locus ab initio, we used the haplotypes

as the initial seeds to pull down long-read Melanesian whole-genome sequence data and iteratively applied the Segmental Duplication Assembler method (41) (fig. S43 and table S14). The procedure generated a ~1.8-Mbp sequence contig spanning more than 900 kbp of complex segmental duplications (Fig. 3A). We confirmed the organization and sequence accuracy (99.86%) for 220 kbp of this assembly for which we have finished BAC sequences derived from an unrelated Melanesian sample (fig. S44). Notably, the sequence-resolved assembly shows that the actual length of DUP_{16p12} duplication polymorphism is ~383 kbp, which is longer than previously thought (9). Sequence and phylogenetic analyses suggest that the variant originated from a series of complex structural changes involving duplication, deletion, and inversion events ~0.5 to 2.5 million years ago (Mya) within the Denisovan ancestral lineage, which subsequently inserted into chromosome 16p11.2 (chr16:29,640,235-29,640,459) between 0.2 to 0.5 million years ago (Fig. 3B, figs. S45 and S46, and table S15).

Within the sequence immediately flanking the 16p11.2 duplication block near the DUP_{16p12} insertion breakpoint, we observe significantly elevated PBS ($P < 0.012$) and f_D ($P < 0.021$, archaic = Denisovan) values (Fig. 2D). Once again, the introgression signal at 16p11.2 is absent when Neanderthals are used as the archaic reference [$P(f_D)$; archaic = Neanderthal] > 0.193; fig. S47], consistent with a Denisovan origin. To infer the timing of the introgression event, we constructed DUP_{16p12} haplotypes for the Denisovan and Melanesian genomes by remapping the short-read data to the assembled Melanesian contig and KV880768.1 in addition to the human reference. We reconstructed the phylogeny of DUP_{16p12} using ~10-kbp sequen-

ces, where at least five high-quality Denisovan reads are present, and showed that the introgression event likely occurred ~0.06 to 0.17 million years ago (fig. S48).

The Melanesian duplication polymorphism harbors extra copies of segmental duplication sequences that are absent from most human populations, including an additional member of the *NPIP* family (42). To explore the *NPIP* coding potential at this locus, we generated full-length nonchimeric (FLNC) transcript data using Melanesian (GM10539 and GM10541) fibroblast cell lines (29). We identified FLNC transcripts that maintain the same open reading frame (ORF) and encode a previously unidentified member of the *NPIP* family, *NPIPBI6* (1206 amino acids), mapping exclusively to this duplication polymorphism (Fig. 3C). This Melanesian copy shows elevated pairwise dN/dS ratios (i.e., the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site) when compared with other closely related *NPIP* genes (RefSeq release 109) (figs. S49 and S50). Using a phylogenetic branch site test (43), we identified 32 sites as likely positively selected, including a cluster of 28 amino acid differences that occur in the last exon of *NPIPBI6* (Fig. 3C). Multiple sequence alignment analyses reveal that this cluster is due to two indel events of a repeat motif (GAG-CGTCTGCGGG). The first indel upstream of the cluster alters the frame, whereas the second indel located downstream subsequently restores the original *NPIP* frame. These compensatory mutations resulted in a novel sequence at the C terminus of *NPIPBI6* (fig. S51), as opposed to the other four amino acid replacement events. Notably, at 30 of the 32 sites where there is sequence coverage (more than five reads), the

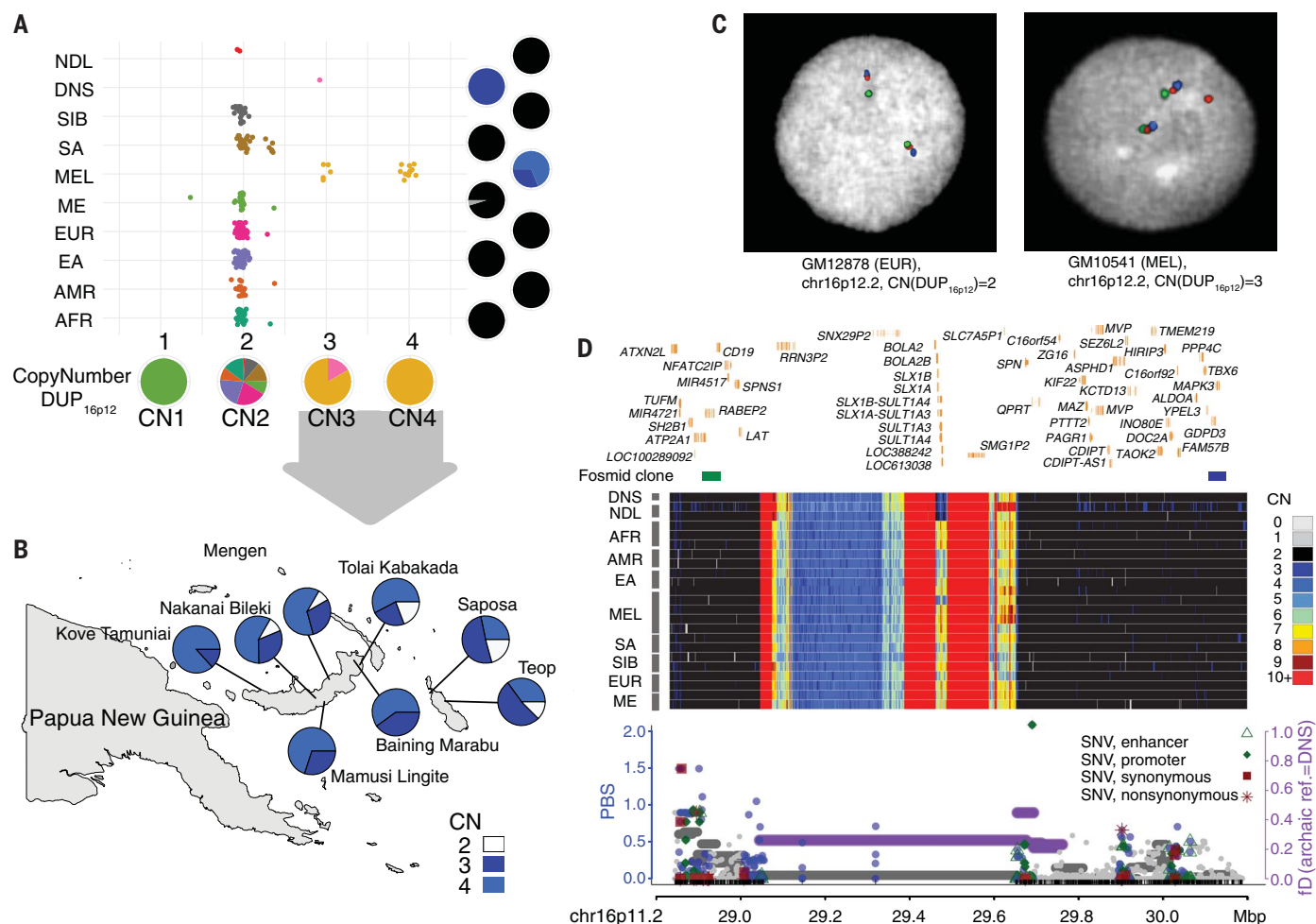


Fig. 2. Evidence for adaptive Denisovan introgression of the chromosome 16p12.2 duplication at 16p11.2 in Melanese. (A) Copy number (CN) estimates of DUP_{16p12} for the SGDP populations and three archaic samples. Pie charts indicate the CN frequency of populations (right) and the population fraction of CN genotypes (bottom). NDL, Neanderthals; DNS, Denisovans; SIB, Siberians; SA, South Asians; MEL, Melanese; ME, Middle Easterners; EUR, Europeans; EA, East Asians; AMR, Native Americans; AFR, sub-Saharan Africans. (B) Geographic distribution for the DUP_{16p12} duplication genotypes of 242 independent blood-derived DNA samples from Melanese. The CN color scheme matches that in (D). (C) FISH experiments using fosmid clones from 16p12.2 confirm an additional copy of DUP_{16p12} (red fosmid clone, 174222_ABC10_2_1_000044550500_M3 at 16p12.2; table S13; fig. S38) in a

Melanese cell line (GM10541, CN3) as opposed to a European cell line (GM12878, CN2). (D) (Top) Signals of adaptive introgression in the Melanese at 16p11.2—the locus in which the DUP_{16p12} duplication was inserted. The heat map shows the CN distribution at chromosome 16p11.2. Fosmid clones (green: ABC10_000044688200_G16; blue: ABC10_000043626100_E12; table S13) indicate the region where the integration of DUP_{16p12} occurred at the 16p11.2 locus. (Bottom) PBS (left y axis) for SNVs (dots) and f_D [horizontal lines, representing windows of 100 SNVs, computed using Denisovans as the archaic reference, right axis] at DUP_{16p12}. Colored circles (blue) and/or horizontal lines (purple) indicate significant test statistics ($P < 0.05$). Note that introgression signals at both 16p12.2 and 16p11.2 disappear if Neanderthals are used as the archaic reference in the f_D computation (figs. S36 and S47).

Denisovan genome carries the same alleles as in the Melanese *NP1PB16*, suggesting that most of these events emerged in the Denisovan lineage.

Taken together, our findings suggest that the 383-kbp duplication polymorphism in Melanese introgressed from Denisovan-like hominins and that it is likely an adaptive CNV. This helps to explain why this polymorphism has become nearly fixed within the Melanese populations (>80%) despite its large size, which is typically regarded as selectively disadvantageous (8, 44). Our analyses support a model where the locus on chromosome 16p11.2 served

as an acceptor site for large independent duplication events twice in the recent history of the *Homo* genus—once in the *H. sapiens* lineage and once in the ancestral Denisovan lineage where novel duplicate genes have emerged. Notably, the Melanese-specific gene *NP1PB* shows ~3% amino acid divergence and evidence of positive selection despite its recent origin. Introgression has generated an amalgam of these two duplications, creating a massive duplication block (~900 kbp) in Melanese and related populations. Given the localization of this duplication polymorphism to the recurrent autism-associated breakpoint region,

we predict that these changes in the genomic architecture and the additional *NP1PB* homologous segments (orange triangles in Fig. 3; see also table S16) will affect the frequency of recurrent rearrangements associated with autism in these populations.

Introgression of Neanderthal CNVs at chromosome 8p21.3 shows signals of positive selection

Another signal of selection (PBS/F_{ST}) and introgression ($P < 0.005$) in Melanese maps to a 76-kbp region on chromosome 8 (chr8:22, 969,611–23,045,069; Fig. 4, A and B, and

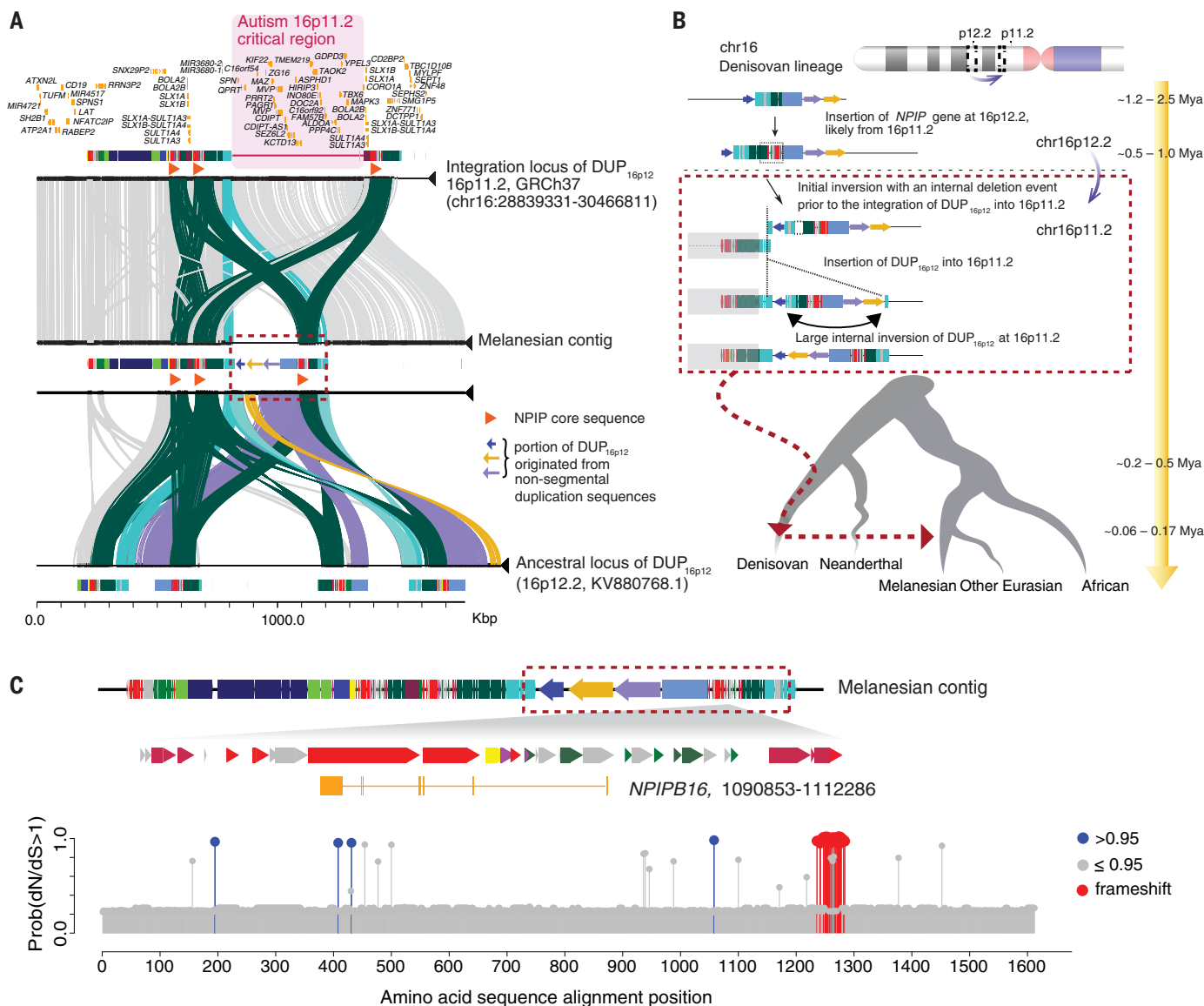


Fig. 3. Reconstruction of the structure and evolutionary history for the Melanesian-Denisovan duplication at chromosome 16p11.2. (A) Structural comparison of chromosome 16p11.2 (human genome reference GRCh37; top), the structure-resolved Melanesian contig (middle) at 16p11.2, and the ancestral locus of DUP_{16p12} at 16p12.2 (KV880768.1; NCBI BioProject: PRJNA31257; bottom). Colored boxes denote annotated human segmental duplications, and lines connecting the sequences show regions of homology. Duplicated segments specific to the Melanesian genome (red dashed box) are indicated if derived from unique (colored arrows) or previous duplication (colored rectangles) sequences. The region of recurrent genome rearrangements associated with autism is highlighted (pink shaded area). (B) Schematic model for the evolution of the DUP_{16p12} duplication. The schematic depicts

structural changes over time, leading to the Melanesian architecture. Evolutionary timing was estimated on the basis of a series of phylogenetic analyses using structure-resolved sequences from 16p12.2 and 16p11.2 loci (31). The absence of intermediate genomes makes the order of some structural changes uncertain. (C) A new member of the *NPIPB* gene family, *NPIPB16* (1206 amino acids), in the Melanesian DUP_{16p12} sequence with predicted sites of positive selection. *dN/dS* analyses show positively selected amino acid substitutions at *NPIPB16* lineages (blue circles) compared with other *NPIPB* genes. Note that the cluster of massive amino acid changes (red circles) at position 1236 to 1284 (alignment space) is predicted to result from two indel events in the C terminus of *NPIPB16* as opposed to a series of independent amino acid substitution events (fig. S51).

fig. S52). Within this region, there are two stratified CNVs: a 6-kbp deletion (DEL_{MEL-NDL}, chr8:22,982,302-22,988,251; CNV stratification $P < 8.9 \times 10^{-11}$) and a 31-kbp duplication (DUP_{10D}, chr8:22,991,347-23,022,738; CNV stratification $P < 1.5 \times 10^{-6}$) encompassing *TNFRSF10D* (tumor necrosis factor receptor superfamily 10D). The copy number estimates

of both structural variants (DEL_{MEL-NDL} and DUP_{10D}) are highly correlated among Melanesian samples (Pearson's correlation $r = -0.64$, $P < 0.05$) but not in other SGDP populations (fig. S53). Multiple lines of evidence, including signals of excess heterozygosity, sequence read-depth analyses, and FISH assays, confirm that the duplication variant DUP_{10D} is the major

allele among Melanesians (AF = 0.66) but is also found among a subset of Africans (AF = 0.12) as well as in the three archaic genomes where it is heterozygous (AF = 0.5) (Fig. 4B and figs. S53 to S55). The deletion allele DEL_{MEL-NDL} shows a restricted distribution, as it is observed among Melanesians (AF = 0.437), the two Neanderthals (AF = 1.0), and at low frequency

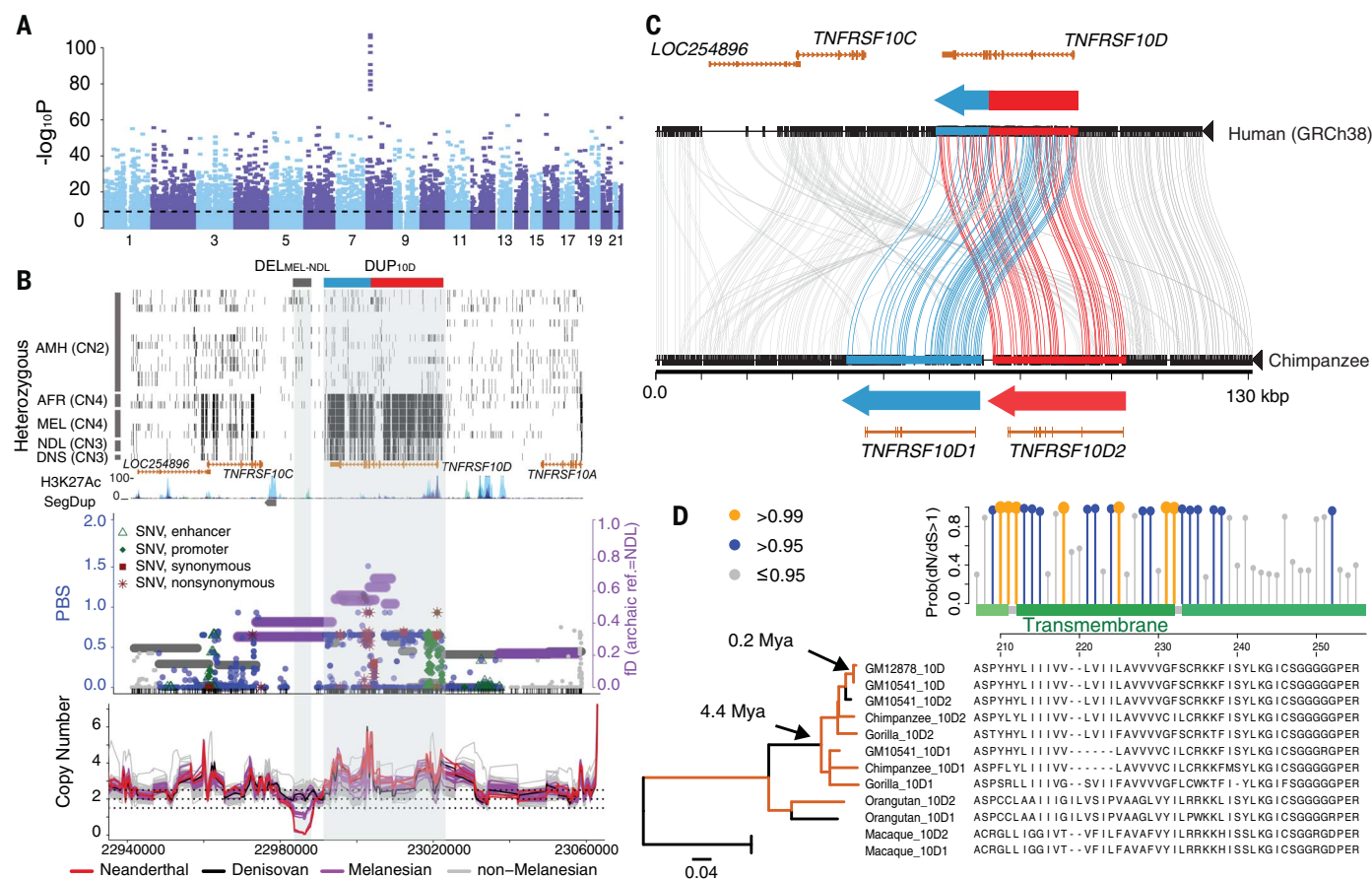


Fig. 4. Highly stratified CNVs at 8p21.3 in Melanesians and evidence for gene duplication and fusion followed by adaptive evolution at the *TNFRSF10D* locus. (A) Manhattan plot for the P values of window-based F_{ST} test. The horizontal dashed line indicates the genome-wide Bonferroni-corrected significance. (B) (Top) Distribution of heterozygous sites (short black vertical bars) for a subset of the SGDP samples. The gray box at the top shows the location of $DEL_{MEL-NDL}$, whereas the blue-red box indicates the derived $TNFRSF10D$ form, a fusion of $TNFRSF10D1$ (blue) and $TNFRSF10D2$ (red), as shown in Fig. 3C. (Bottom) Distributions of f_D and PBS statistics, as well as CN trajectories of all samples

across the region. (C) Comparison (Miropeats) of the major human allele versus chimpanzee genome structure, showing the tandem organization of the DUP_{10D} variant and the predicted gene models. (D) Branch-site test of positive selection (dN/dS) using FLNC transcript data shows significant selection signals ($P = 0.005$) compared with the null model and a cluster of positively selected amino acid substitutions at the transmembrane domain of $TNFRSF10D$. Coding-sequence phylogeny shows significant positive selection (orange; dN/dS ratios > 1 ; $P < 0.05$) on specific branches. Note that the orangutan paralogous sequences form a single clade as a result of interlocus gene conversion (fig. S60).

among South Asians ($AF = 0.013$). Using a PCR assay, we confirmed the presence of the $DEL_{MEL-NDL}$ deletion allele in two independent sets of cell line and blood-derived Papuan DNA samples (fig. S56 and table S12) (29).

To understand the evolution of DUP_{10D} , we computed its copy number among great apes (28) and samples from the 1000 Genomes Project (1KG) (10). The DUP_{10D} variant is present among all nonhuman primates examined and segregates at low frequencies across the 1KG populations (<0.025) but is completely absent from others, such as the Europeans (figs. S53 to S55 and S57). We used BACs to sequence the DUP_{10D} locus from chimpanzees, gorillas, orangutans, and rhesus macaques (29). Our sequence comparison analyses showed that all nonhuman ape species share the same tandem organization for the two copies of $TNFRSF10D$ ($TNFRSF10D1$ of 30.3 kbp and

$TNFRSF10D2$ of 33.0 kbp) (Fig. 4C and fig. S58). The absence of the tandem structure from the human reference genome and most extant humans suggests that the single-copy structure of $TNFRSF10D$ emerged specifically in the human lineage.

Using the chimpanzee sequence, we inferred the evolutionary breakpoints at chr8:23,003,123-23,003,255 (GRCh37) partially overlapping the fifth intron and exon of $TNFRSF10D$. We predict that the most common derived form of $TNFRSF10D$ in humans is a product of a nonallelic homologous recombination event between two ancestral genes ($TNFRSF10D1$ and $TNFRSF10D2$). The resulting gene fusion creates a hybrid consisting of the first five exons of $TNFRSF10D2$ and the last four exons of $TNFRSF10D1$. Thus, two-thirds of $TNFRSF10D1$ (~18.4 kbp) and one-third of $TNFRSF10D2$ (~11.9 kbp) are absent from most humans (figs.

S58 and S59). Bayesian phylogenetic reconstruction of the loci show that two independent duplication events of $TNFRSF10D$ occurred in primate evolution: one in the lineage leading to the Old World monkey and the other in the ancestral lineage of all apes ~27.6 million years ago (95% highest posterior density: 19.9 to 36.1 million years ago; fig. S60). In contrast, we estimate that the gene fusion occurred 1.1 to 1.5 million years ago in the ancestral *Homo* lineage leading to modern humans, Neanderthals, and Denisovans, where it rose to high frequency among most human populations, with the exception of Melanesians.

We assessed the coding potential of the three different $TNFRSF10D$ models by sequencing full-length transcripts from Melanesian (GM10541, CN3), European (GM12878, CN2), and chimpanzee (PanTro, CN4) cell line samples (29). In chimpanzee cells, we identified a six- and a nine-exon

gene model corresponding to *TNFRSF10D1* and *TNFRSF10D2*, respectively (figs. S61 and S62). In contrast, the European CN2 sample possesses only the fusion gene, *TNFRSF10D*, whereas the heterozygous Melanesian sample shows all three gene models, including the two additional models present in chimpanzee (fig. S61). Although ORF analyses indicate that *TNFRSF10D1* carries a premature stop codon in exon 2, which truncates the canonical protein at the 59th amino acid position, a protein with 217 amino acids is still predicted, owing to the use of a second start codon in exon 2 (fig. S62). In contrast, *TNFRSF10D2* maintains an ORF with all nine exons consistent with the annotation in the human reference assembly (ENST00000312584). The more common human fusion gene effectively deletes the *TNFRSF10D1* premature stop codon (fig. S62) and thus restores the canonical ORF.

To further assess the functional importance of the *TNFRSF10D* duplicates, we examined FLNC transcript data from Melanesian (GM10541), European (GM12878), and chimpanzee fibroblast cell lines (29) and inferred orthologous gene models from *TNFRSF10D1* and *TNFRSF10D2* using gorilla, orangutan, and rhesus macaque BAC sequences. We observed a significant excess of amino acid substitutions for the three gene models, especially among the African great ape lineages (fig. S63) (29). Using a branch-site test (43), we find evidence of positive selection for both the fusion gene and the *TNFRSF10D1* copy on the human lineage and, more broadly, for both D1 and D2 in other nonhuman apes ($P = 0.005$; Fig. 4D and figs. S64 and S65). Specifically, we identify a cluster of positively selected sites corresponding to the predicted transmembrane domain of the genes (Fig. 4D and fig. S64).

In contrast to the *TNFRSF10D* duplication, the DEL_{MEL-NDL} variant is absent from all nonhuman apes, suggesting that it is derived (fig. S53). Using a PCR assay (table S17) and seven tagging SNVs [$r^2 > 0.9$ and normalized coefficient of linkage disequilibrium (D') > 0.9 ; table S18], we show that the DEL_{MEL-NDL} variant is largely restricted to the 1KG South Asian populations where it occurs at low frequencies (<0.07 ; Fig. 5A and figs. S57 and S66). All three Neanderthal genomes, including the low-coverage Mezmaiskaya (24), are homozygous for the DEL_{MEL-NDL} haplotype, suggesting that this deletion is fixed in Neanderthals (table S18). We identify strong Neanderthal introgression signals at DEL_{MEL-NDL} among Melanesians [$P(f_{\text{D}}$, archaic = Neanderthal) = 0.003 and $P(S^*) = 0.043$; Fig. 4B and fig. S67], as opposed to Denisovan introgression [$P(f_{\text{D}}$, archaic = Denisovan) = 0.06; fig. S68]. Consistent with this, we find that all 15 DEL_{MEL-NDL}-linked SGDP haplotypes are more closely related to the Neanderthal haplotypes than any other human haplotype (fig. S69), forming

a monophyletic clade in both the haplotype network (Fig. 5B) and maximum likelihood estimated phylogenetic tree (log-likelihood = $-21,578$; fig. S70) (31). The time to the most recent common ancestor of the Neanderthal and Melanesian DEL_{MEL-NDL}-linked haplotypes is 40,000 years ago (95% CI: 0 to 122,000 years ago) and that of all human DEL_{MEL-NDL}-linked haplotypes is 120,000 years ago (95% CI: 0 to 241,000 years ago)—both are much younger than the divergence of modern humans and Neanderthals. It is worth noting that the haplotype sharing of the 18.5-kbp sequence between modern humans and Neanderthals is unlikely under a simple model of incomplete lineage sorting (ILS) with reasonable demographic parameters ($P = 0.0403$) (29).

Because the large frequency difference of the DEL_{MEL-NDL} haplotype between Melanesians and other populations can be due to the action of natural selection, we specifically tested this hypothesis. Using the deletion variant as a proxy for the beneficial allele and controlling the age of the variant in coalescent simulations (29), we find that the observed PBS value of the DEL_{MEL-NDL} deletion allele is significantly higher than expected under a null model of neutral evolution (PBS = 0.933, $P = 0.0082$; Fig. 5C). Among Melanesians, the deletion allele is always associated with DUP_{10D}. This 6-kbp deletion corresponds to a histone H3 at lysine 27 (H3K27) acetylation peak (skeletal muscle tissue), which by ChIA-PET (chromatin interaction analysis by paired-end tag) analysis associates with multiple members of the tumor necrosis factor receptor superfamily—a gene family critical for apoptosis (45).

From these observations, we developed a model to account for the recent evolution of this 8p21.3 locus (Fig. 5D). Our data support a copy number reduction of *TNFRSF10D* to create a fusion gene approximately 1.2 to 1.5 million years ago in the ancestral *Homo* lineage. The presence of the ancestral *TNFRSF10D1* and *TNFRSF10D2* in some, but not all, modern human populations likely emerged as a result of ILS. A 6-kbp deletion (DEL_{MEL-NDL}) located ~5 kbp distal to the *TNFRSF10D* locus in a putative regulatory region subsequently originated in the Neanderthal lineage, which then introgressed back into the human lineage <120,000 years ago through contact with ancestral non-Africans. We predict that the unusually high frequency of the DEL_{MEL-NDL}-linked haplotype in Melanesians is a result of natural selection. The ChIA-PET data suggest that this deletion may affect the expression of nearby genes, such as *TNFRSF10C* and *TNFRSF10D* (WashU Epigenome Browser: <https://epigenomegateway.wustl.edu>).

Discussion

Despite the growing body of evidence for adaptive introgression in humans (5–7), the ex-

tent and relative contributions of adaptive introgressed CNVs in human evolution remain largely unexplored. Leveraging sequences from archaic and modern human genomes identifies adaptive introgressed CNV loci, which we resolved and characterized. The enrichment of selective and adaptive introgressed CNVs highlights their potential role in human evolution. Although we cannot rule out potential “hitchhiking” effects due to positive selection acting on nearby genes, the lack of other functional mutations of large effect (i.e., nonsynonymous variants) around our candidate loci suggests that these selection signals are driven by the stratified CNVs identified here and possibly by the genes carried within.

One possible confounding factor in such an analysis is ILS; however, we believe that this has had limited effect for several reasons. First, the time to the most recent common ancestor estimates between the Melanesian and archaic haplotypes at the 16p11.2 and 8p21.3 loci is less than 170,000 years, which is inconsistent with ILS. Second, we account for Melanesian demographic history in determining thresholds of significance for the individual population genetic tests applied here. Finally, tests of positive selection using dN/dS are generally unaffected by ILS (46, 47) and thus cannot account for the signals of positive selection we observe. It should be noted that SNVs that tag the CNV genotypes at the candidate loci of selection in Melanesians are generally specific to Melanesians, with relatively little association in other SGDP populations (9). For example, in the case of the *APOEC3A* and *APOEC3B* deletion variant, we find 266 SNVs from the 110-kbp region in almost perfect linkage disequilibrium with the deletion allele in Melanesians, but among other East Asian populations the association is weaker (fig. S19).

Although our limited understanding of genotype–phenotype relationships makes it difficult to predict functional consequences of the CNVs at chromosomes 16p11.2 and 8p21.3, we note that chromosome 16p11.2 harbors adaptive signals associated with human-specific gene duplication expansion that may affect iron regulation in human embryonic development (15). This locus also exhibits an enrichment of complex recurrent structural rearrangements, which predisposes humans to the second most common genetic cause of autism, accounting for ~1% of patients (15, 40). In the case of the 8p21.3 CNVs, the Melanesian DUP_{10D} allele is accompanied by a nearby deletion (DEL_{MEL-NDL}) that originated in Neanderthals and overlaps with known regulatory signals in human skeletal muscle myoblasts (H3K27Ac markers, UCSC Genome Browser). Notably, *TNFRSF10D* has been reported as one of the primate-specific genes preferentially expressed in progenitor cells of the human fetal neocortex (48).

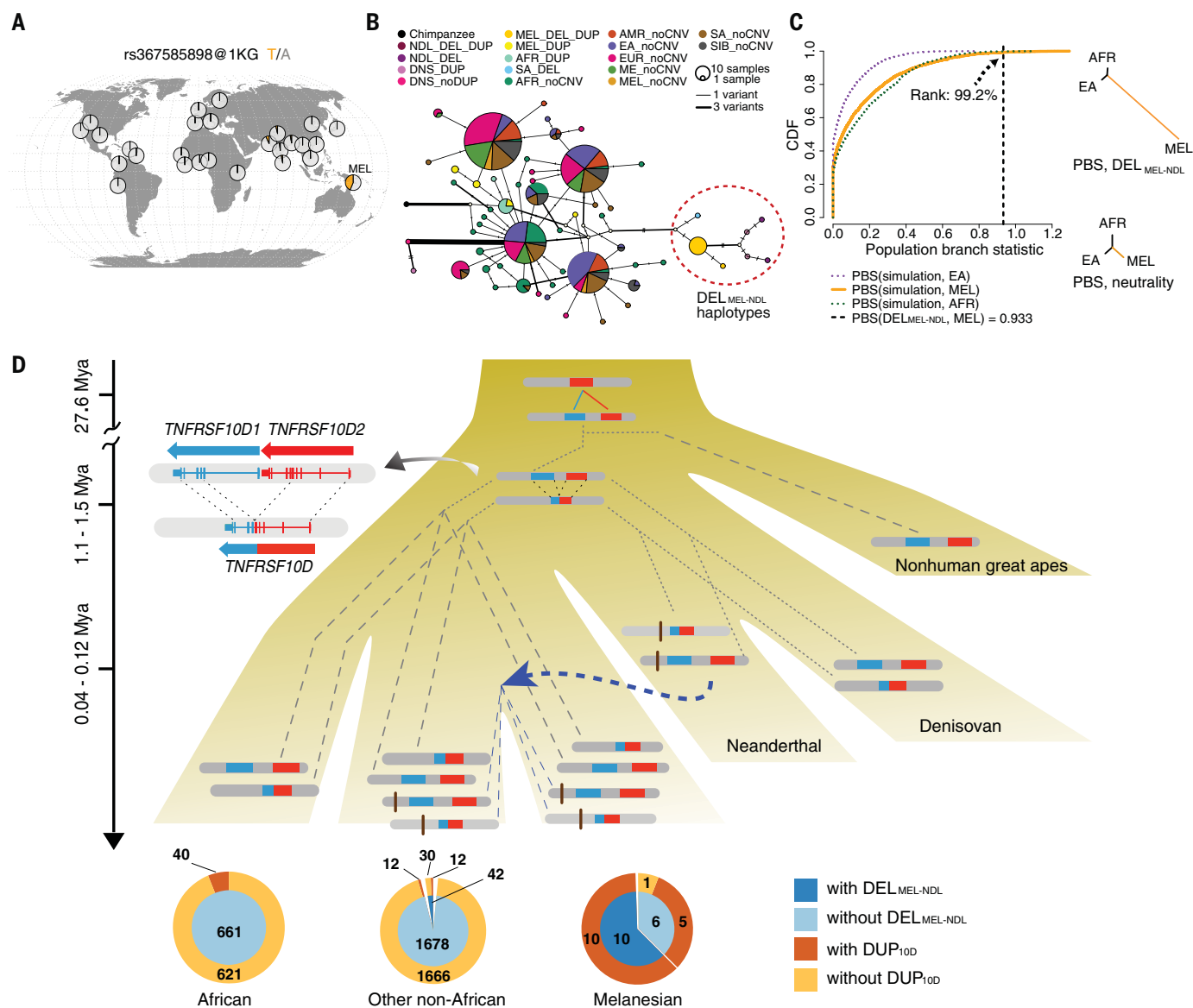


Fig. 5. Complex evolutionary history and evidence of adaptive archaic introgression at 8p21.3 in Melanesians. (A) Geographic distribution of the DEL_{MEL-NDL} tag SNV, rs367585898, using the 1KG data. (B) Median-joint network showing that all deletion-linked haplotypes form a single clade. Colors represent different haplotypes (DEL and DUP refer to DEL_{MEL-NDL} and DUP_{10D}, respectively). Edge width indicates the number of pairwise differences between the joined haplotypes. (C) A highly unusual PBS of DEL_{MEL-NDL} in the Melanesians,

compared with those computed using variants with similar age in neutral coalescent simulations, supports the hypothesis of positive selection favoring this introgressed variant. CDF, cumulative distribution function. (D) (Top) Schematic model for the evolution of the DEL_{MEL-NDL} and DUP_{10D} variants among great ape lineages. Dashed and dotted lines correspond to contemporary and historical lineages, respectively. (Bottom) Pie charts show the numbers of individuals from the 1KG and SGDP panels with or without the two CNVs.

Large genomic gains and losses have the potential to create genes that can affect organismal phenotypes. With our increasing understanding of gene flow between various archaic forms and our ancestors, it is tempting to hypothesize that CNV introgression from other hominins may have played a key role in helping humans migrating out of Africa adapt to new environments by serving as a reservoir of beneficial alleles. In addition, this study highlights the substantial large-scale genetic variation that remains to be characterized in the human population and the need for develop-

ment of additional reference genomes that better capture the diversity of our species and complete our understanding of human genes (11).

Materials and methods summary

We downloaded paired-end Illumina data for 266 fully publicly available human genomes via the SGDP website (27) and called SNVs and indels using the HaplotypeCaller function in Genome Analysis Toolkit (GATK, v3.5). After quality control filtering, these variants were merged with genotypes from the three archaic

genomes (<http://cdna.eva.mpg.de/>) using BCFtools (v1.5). For the SGDP samples, CNV calls were generated using WHAMG, LUMPY (v0.2.13), DELLY2 (v0.7.2), digital comparative genomic hybridization (dCGH), and Genome STRiP (v2.00.1611). We applied only dCGH to the three archaic genomes, owing to the lack of paired-end data in these genomes. Population-stratified CNVs were identified using V_{ST} , MWU, and D_{median} statistics, whereas signals of positive selection and archaic introgression were determined using the PBS, f_D , and S^* statistics (2, 30, 49). Significance of these test

results was determined from coalescent simulations based on the best-fit demographic models inferred from *∂a∂i* (50) and ADMIXTURE (51). Phylogenetic analyses were performed with a combination of RAxML (v8.2.10), BEAST (v2.5.0), and PAML (v1.4.9). We used FISH, sequencing of BAC clone inserts, and long-read sequencing data to validate CNVs and assess the breakpoints of complex rearrangements. The high-coverage long-read sequence data from the Melanesian sample was generated using PacBio single-molecule, real-time (SMRT) sequencing technology. We followed the Iso-Seq protocol and used PacBio sequencing (52) to generate the FLNC transcript data. Detailed descriptions of materials and methods are available in the supplementary materials (29).

REFERENCES AND NOTES

1. S. A. Tishkoff *et al.*, Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007). doi: [10.1038/ng1946](#); pmid: [17159977](#)
2. X. Yi *et al.*, Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010). doi: [10.1126/science.1190371](#); pmid: [20595611](#)
3. M. Fumagalli *et al.*, Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* **349**, 1343–1347 (2015). doi: [10.1126/science.aab2319](#); pmid: [26383953](#)
4. M. A. Ilardo *et al.*, Physiological and genetic adaptations to diving in sea nomads. *Cell* **173**, 569–580.e15 (2018). doi: [10.1016/j.cell.2018.03.054](#); pmid: [29677510](#)
5. F. L. Mendez, J. C. Watkins, M. F. Hammer, A haplotype at STAT2 introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am. J. Hum. Genet.* **91**, 265–274 (2012). doi: [10.1016/j.ajhg.2012.06.015](#); pmid: [22883142](#)
6. E. Huerta-Sánchez *et al.*, Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014). doi: [10.1038/nature13408](#); pmid: [25043035](#)
7. F. Racimo *et al.*, Archaic Adaptive Introgression in TBX15/WARS2. *Mol. Biol. Evol.* **34**, 509–524 (2017). pmid: [28007980](#)
8. D. F. Conrad *et al.*, Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010). doi: [10.1038/nature08516](#); pmid: [19812545](#)
9. P. H. Sudmant *et al.*, Global diversity, population stratification, and selection of human copy-number variation. *Science* **349**, aab3761 (2015). doi: [10.1126/science.aab3761](#); pmid: [26249230](#)
10. P. H. Sudmant *et al.*, An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015). doi: [10.1038/nature15394](#); pmid: [26432246](#)
11. P. A. Audano *et al.*, Characterizing the major structural variant alleles of the human genome. *Cell* **176**, 663–675.e19 (2019). doi: [10.1016/j.cell.2018.12.019](#); pmid: [30661756](#)
12. G. H. Perry *et al.*, Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007). doi: [10.1038/ng2123](#); pmid: [17828263](#)
13. Y. Xue *et al.*, Adaptive evolution of UGT2B17 copy-number variation. *Am. J. Hum. Genet.* **83**, 337–346 (2008). doi: [10.1016/j.ajhg.2008.08.004](#); pmid: [18760392](#)
14. R. J. Hardwick *et al.*, A worldwide analysis of beta-defensin copy number variation suggests recent selection of a high-expressing DEFB103 gene copy in East Asia. *Hum. Mutat.* **32**, 743–750 (2011). doi: [10.1002/humu.21491](#); pmid: [21387465](#)
15. X. Nettle *et al.*, Emergence of a Homo sapiens-specific gene family and chromosome 16p11.2 CNV susceptibility. *Nature* **536**, 205–209 (2016). doi: [10.1038/nature19075](#); pmid: [27487209](#)
16. S. Lindeberg, P. Nilsson-Ehle, B. Vessby, Lipoprotein composition and serum cholesterol ester fatty acids in nonwesternized Melanesians. *Lipids* **31**, 153–158 (1996). doi: [10.1007/BF02522614](#); pmid: [8835402](#)
17. J. Flint *et al.*, High frequencies of alpha-thalassaemia are the result of natural selection for malaria. *Nature* **321**, 744–750 (1986). doi: [10.1038/321744a0](#); pmid: [3713863](#)
18. P. T. Katzmarzyk, W. R. Leonard, Climatic influences on human body size and proportions: Ecological adaptations and secular trends. *Am. J. Phys. Anthropol.* **106**, 483–503 (1998). doi: [10.1002/\(SICI\)1096-8644\(199808\)106:4<483::AID-AJPA4>3.0.CO;2-K](#); pmid: [9712477](#)
19. A. S. Malaspina *et al.*, A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016). doi: [10.1038/nature18299](#); pmid: [27654914](#)
20. A. Bergström *et al.*, A Neolithic expansion, but strong genetic structure, in the independent history of New Guinea. *Science* **357**, 1160–1163 (2017). doi: [10.1126/science.aan3842](#); pmid: [28912245](#)
21. D. Reich *et al.*, Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010). doi: [10.1038/nature09710](#); pmid: [21179161](#)
22. P. Skoglund *et al.*, Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510–513 (2016). doi: [10.1038/nature19844](#); pmid: [27698418](#)
23. B. Vernot *et al.*, Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016). doi: [10.1126/science.aad9416](#); pmid: [26989198](#)
24. K. Prüfer *et al.*, A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**, 655–658 (2017). doi: [10.1126/science.aao1887](#); pmid: [28982794](#)
25. M. Meyer *et al.*, A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012). doi: [10.1126/science.1224344](#); pmid: [22936568](#)
26. K. Prüfer *et al.*, The complete genome sequence of a Neandertal from the Altai Mountains. *Nature* **505**, 43–49 (2014). doi: [10.1038/nature12886](#); pmid: [24352235](#)
27. S. Mallick *et al.*, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi: [10.1038/nature18964](#); pmid: [27654912](#)
28. J. Prado-Martinez *et al.*, Great ape genetic diversity and population history. *Nature* **499**, 471–475 (2013). doi: [10.1038/nature12228](#); pmid: [23823723](#)
29. Materials and methods are available as supplementary materials.
30. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015). doi: [10.1093/molbev/msu269](#); pmid: [25246699](#)
31. V. Tillander, S. E. H. Alexson, D. E. Cohen, Deactivating Fatty Acids: Acyl-CoA Thioesterase-Mediated Control of Lipid Metabolism. *Trends Endocrinol. Metab.* **28**, 473–484 (2017). doi: [10.1016/j.tem.2017.03.001](#); pmid: [28385385](#)
32. J. M. Kidd, T. L. Newman, E. Tuzun, R. Kaul, E. E. Eichler, Population stratification of a common APOBEC gene deletion polymorphism. *PLOS Genet.* **3**, e63 (2007). doi: [10.1371/journal.pgen.0030063](#); pmid: [17447845](#)
33. P. An *et al.*, APOBEC3B deletion and risk of HIV-1 acquisition. *J. Infect. Dis.* **200**, 1054–1058 (2009). doi: [10.1086/605644](#); pmid: [19698078](#)
34. N. J. Smith, T. R. Fenton, The APOBEC3 genes and their role in cancer: Insights from human papillomavirus. *J. Mol. Endocrinol.* **62**, R269–R287 (2019). doi: [10.1530/JME-19-0011](#); pmid: [30870810](#)
35. Y. Y. Qi *et al.*, DEFA gene variants associated with IgA nephropathy in a Chinese population. *Genes Immun.* **16**, 231–237 (2015). doi: [10.1038/gene.2015.1](#); pmid: [25675412](#)
36. K. Mohajeri *et al.*, Interchromosomal core duplicons drive both evolutionary instability and disease susceptibility of the Chromosome 8p23.1 region. *Genome Res.* **26**, 1453–1467 (2016). doi: [10.1101/gr.211284.116](#); pmid: [27803192](#)
37. K. M. Steinberg *et al.*, Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat. Genet.* **44**, 872–880 (2012). doi: [10.1038/ng.2335](#); pmid: [22751100](#)
38. H. Stefansson *et al.*, A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005). doi: [10.1038/ng1508](#); pmid: [15654335](#)
39. F. Antonacci *et al.*, A large and complex structural polymorphism at 16p12.1 underlies microdeletion disease risk. *Nat. Genet.* **42**, 745–750 (2010). doi: [10.1038/ng.643](#); pmid: [20729854](#)
40. L. A. Weiss *et al.*, Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008). doi: [10.1056/NEJMoa075974](#); pmid: [18184952](#)
41. M. R. Vollger *et al.*, Long-read sequence and assembly of segmental duplications. *Nat. Methods* **16**, 88–94 (2019). doi: [10.1038/s41592-018-0236-3](#); pmid: [30559433](#)
42. M. E. Johnson *et al.*, Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**, 514–519 (2001). doi: [10.1038/35097067](#); pmid: [11586358](#)
43. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007). doi: [10.1093/molbev/msm088](#); pmid: [17483113](#)
44. B. P. Coe *et al.*, Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat. Genet.* **51**, 106–116 (2019). doi: [10.1038/s41588-018-0288-4](#); pmid: [30559488](#)
45. G. Pan, J. Ni, G. Yu, Y. F. Wei, V. M. Dixit, TRUND, a new member of the TRAIL receptor family that antagonizes TRAIL signalling. *FEBS Lett.* **424**, 41–45 (1998). doi: [10.1016/S0014-5793\(98\)00135-5](#); pmid: [9537512](#)
46. A. Scally *et al.*, Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169–175 (2012). doi: [10.1038/nature10842](#); pmid: [22398555](#)
47. D. A. Pollard, V. N. Iyer, A. M. Moses, M. B. Eisen, Widespread discordance of gene trees with species tree in *Drosophila*: Evidence for incomplete lineage sorting. *PLOS Genet.* **2**, e173 (2006). doi: [10.1371/journal.pgen.0020173](#); pmid: [17132051](#)
48. M. Florio *et al.*, Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. *eLife* **7**, e32332 (2018). doi: [10.7554/eLife.32332](#); pmid: [29561261](#)
49. V. Plagnol, J. D. Wall, Possible ancestral structure in human populations. *PLOS Genet.* **2**, e105 (2006). doi: [10.1371/journal.pgen.0020105](#); pmid: [16895447](#)
50. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genet.* **5**, e1000695 (2009). doi: [10.1371/journal.pgen.1000695](#); pmid: [19851460](#)
51. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009). doi: [10.1101/gr.094052.109](#); pmid: [19648217](#)
52. M. L. Dougherty *et al.*, Transcriptional fates of human-specific segmental duplications in brain. *Genome Res.* **28**, 1566–1576 (2018). doi: [10.1101/gr.237610.118](#); pmid: [30228200](#)

ACKNOWLEDGMENTS

We thank J. Friedlaender, M. Brilliant, and T. Carter for providing access to the Papua New Guinea DNA samples housed at the Marshfield Clinic Research Institute in the Center for Precision Medicine Research; J. Akey for providing additional Papuan DNA samples; T. Brown for assistance in editing this manuscript; J. Friedlaender for helpful comments in the preparation of this manuscript; and D. S. Gordon for help with submitting our data to the NCBI database. **Funding:** This work was supported, in part, by U.S. National Institutes of Health (NIH) grant R01HG002385 to E.E.E. S.C. was supported by a National Health and Medical Research Council (NHMRC) C. J. Martin Biomedical Fellowship (1073726). E.E.E. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** P.H., M.R.V., Z.N.K., J.G.U., and E.E.E. designed and planned experiments. V.D., C.B., S.C., K.H., A.P.L., K.M.M., M.S., and J.G.U. prepared libraries and generated and analyzed sequencing data. P.H., M.R.V., V.D., Z.N.K., S.M., and B.J.N. performed variant calling and bioinformatics analyses. P.H., M.R.V., and D.P. analyzed long-read sequencing data and assembled contigs. P.H. performed population genetic and phylogenetic analyses. G.C., F.A.M.M., and F.A. generated and analyzed FISH experiment data. A.P.L., K.M.M., and J.G.U. generated Iso-Seq transcript data. K.H. performed PCR assays for CNV validations. H.B. and J.-F.D. provided Melanesian genome DNA materials. P.H. and E.E.E. wrote the manuscript. **Competing interests:** E.E.E. is on the scientific advisory board of DanaAnexus, Inc. J.G.U. is an employee of Pacific Biosciences, Inc. **Data and material availability:** All data used in this study, including long-read genomic data (HGDP00550), assembled BAC contigs, and Iso-Seq capture transcript data, are deposited in NCBI under BioProject ID PRJNA522307; these data are available to anyone for the purposes of reproducing or extending the analysis. The study design and research protocol were reviewed by Institutional Review Board (49910) at the University of Washington School of Medicine.

SUPPLEMENTARY MATERIALS

[science.sciencemag.org/content/366/6463/eaax2083/suppl/DC1](#)
Materials and Methods
Supplementary Text
Figs. S1 to S71
Tables S1 to S21
References (53–80)

View/request a protocol for this paper from Bio-protocol.

3 March 2019; resubmitted 5 July 2019
Accepted 12 September 2019
[10.1126/science.aax2083](#)

RESEARCH ARTICLE SUMMARY

NEURODEVELOPMENT

Coordination between stochastic and deterministic specification in the *Drosophila* visual system

Maximilien Courgeon and Claude Desplan*

INTRODUCTION: Neuronal fate decisions must be coordinated across brain regions so that the proper stoichiometry of distinct afferents and their target neurons is established. Many sensory systems rely on stochastic fate specification to increase their repertoire of neuronal types and/or to establish the mosaic of distinct sensory receptors. Once these stochastic fate decisions have been made at the periphery, they must be conveyed to processing centers in the brain to ensure that the correct information is retained. This in turn requires that the proper number of target neurons of the proper type are generated and are connected to their cognate sensory receptors. How these stochastically de-

termined sensory systems transmit their decision to their targets in the brain—as when red (L) and green (M) cone photoreceptors connect to their appropriate bipolar cells in the human retina, or when randomly specified olfactory receptors project to their correct glomeruli in the mouse olfactory bulb—has remained unknown.

RATIONALE: In the *Drosophila* retina, different types of color-sensitive R7 photoreceptors with different ultraviolet spectral sensitivities are stochastically specified and distributed. By contrast, neurogenesis of the optic lobes that receive input from photoreceptors is highly deterministic: A fixed number of neuroblasts produces se-

quentially the same types of neurons following a rigid temporal program as well as a spatial program of determination. We asked how the stochastic choices made in the retina are propagated to the optic lobes—that is, how stochastically specified photoreceptors project and connect to their neuronal targets that are generated through a highly deterministic program.

RESULTS: We identified distinct subtypes of the main postsynaptic target of the ultraviolet-sensitive R7 photoreceptors, the Dm8 neurons. Each Dm8 subtype is specific to one of the three different subtypes of R7, and their organization in the optic lobe mirrors the topographic organization of R7s in the retina. The three subtypes of Dm8 are prespecified early in neurogenesis by distinct neural progenitors, independent-

ly from (and before) any photoreceptor input. These neurons are produced in excess, but only when the correct photoreceptor-Dm8 pairs form do Dm8s survive. This allows them

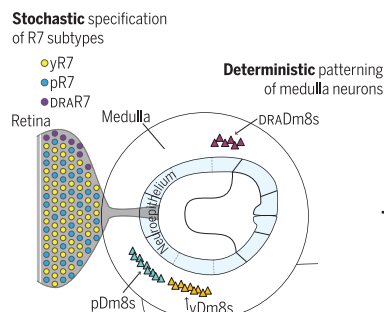
to accommodate random variations in the production of the subtypes of R7s. Those Dm8s that do not find their correct R7 partner are culled by programmed cell death. We then showed that matching of one of the R7-Dm8 subtypes is mediated by two complementary cell adhesion molecules: Dpr11 in one of the subtypes of R7, and its cognate receptor DIP γ in the corresponding subclass of Dm8. Loss of either molecule leads to the death of that Dm8 subtype or its mispairing with the wrong R7 subtype; these phenotypes are similar to the loss of the corresponding subtype of R7 photoreceptors. These mechanisms allow the qualitative and quantitative matching of R7 subtypes with their target in the brain, and thus permit the stochastic choice made in R7 to be propagated to the deterministically specified downstream circuit to support color vision.

CONCLUSION: Our data reveal two fundamental principles in circuit formation: (i) The neurotrophic theory suggests that neuronal survival depends on trophic factors provided by the targets. We show that apoptosis is also involved in the quantitative and qualitative matching of target neurons with their afferent sensory receptors, providing a direct link between correct synaptic partner choice and cell survival. (ii) As Sperry's chemoaffinity theory proposed, we show that synaptic partner choice is controlled by two interacting cell adhesion molecules, which are part of two large families of interacting proteins. ■

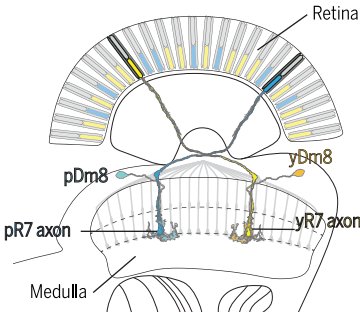
Department of Biology, New York University, New York, NY 10003, USA.

*Corresponding author. Email: cd38@nyu.edu
Cite this article as M. Courgeon, C. Desplan, *Science* 366, eaay6727 (2019). DOI: 10.1126/science.aay6727

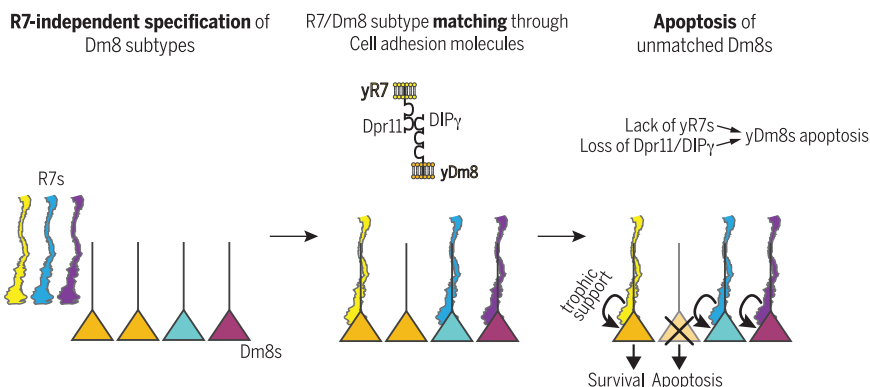
A *Drosophila* visual system development



B Adult visual system



C



Circuit development in the *Drosophila* visual system. (A) During larval development, distinct subtypes of R7 photoreceptors are stochastically specified, whereas medulla neurons are specified through a deterministic program. (B) In the adult optic lobe, every R7 is connected to a Dm8 neuron of the proper subtype. (C) Three steps allow the propagation of the stochastic choice made by R7 in the retina to the optic lobe: Distinct Dm8 subtypes are prespecified and produced in excess independently of R7s. Then, R7-Dm8 pairs of the correct subtype are matched, controlled by the complementary cell adhesion molecules Dpr11 and DIP γ for the yR7-yDm8 pair. Finally, unmatched Dm8s are eliminated by apoptosis.

RESEARCH ARTICLE

NEURODEVELOPMENT

Coordination between stochastic and deterministic specification in the *Drosophila* visual system

Maximilien Courgeon and Claude Desplan*

Sensory systems use stochastic fate specification to increase their repertoire of neuronal types. How these stochastic decisions are coordinated with the development of their targets is unknown. In the *Drosophila* retina, two subtypes of ultraviolet-sensitive R7 photoreceptors are stochastically specified. In contrast, their targets in the brain are specified through a deterministic program. We identified subtypes of the main target of R7, the Dm8 neurons, each specific to the different subtypes of R7s. Dm8 subtypes are produced in excess by distinct neuronal progenitors, independently from R7. After matching with their cognate R7, supernumerary Dm8s are eliminated by apoptosis. Two interacting cell adhesion molecules, Dpr11 and DIP γ , are essential for the matching of one of the synaptic pairs. These mechanisms allow the qualitative and quantitative matching of R7 and Dm8 and thereby permit the stochastic choice made in R7 to propagate to the brain.

Stochastic specification of neurons is a common feature of many sensory systems (1). In the vertebrate olfactory system, it is used to increase the diversity of olfactory sensory neuron types to a repertoire of more than 1400 in mice (2). In humans and old world monkeys, the stochastic specification of cone cells is the basis of the retinal mosaic responsible for trichromatic color vision (3, 4). A neuron that relies on an initial stochastic decision must stabilize its choice to maintain its proper identity and then inform its downstream target cells of its choice. The latter is essential for neurons because they need to connect to their proper targets to faithfully transmit information to processing centers. The mouse olfactory system offers the most noteworthy illustration of this matching problem: The ~1400 olfactory neuron subtypes are randomly distributed within the olfactory epithelium (5), yet all olfactory neurons of the same subtype project to the exact same glomeruli of the olfactory bulb (6–8).

In the *Drosophila* retina, a similar stochastic mechanism ensures the random distribution of photoreceptors with different spectral sensitivity (9, 10). The *Drosophila* compound eye is composed of ~750 unit eyes called ommatidia, each composed of eight photoreceptors. Ommatidia can be classified into different subtypes according to the rhodopsins with different spectral sensitivity expressed by color photoreceptors R7 and R8. The main part of the retina is occupied by two types of ommatidia that are randomly distributed and

stochastically specified [Fig. 1A; reviewed in (11)]. In the yellow (y) type that represents 65% of ommatidia, R7 expresses the ultraviolet (UV)-sensitive Rh4, whereas R8—located below R7, and thus seeing the same point in space—always expresses the green-sensitive Rh6. In the remaining 35% of ommatidia of the pale (p) subtype, R7 expresses the shorter UV-sensitive Rh3 and R8 expresses the blue-sensitive Rh5. A third type of ommatidia called dorsal rim area (DRA) is localized in the most dorsal row of ommatidia (12). In this subtype, both R7 and R8 express Rh3 and are responsible for detecting the polarized skylight used for navigation (13).

Most of the gene regulatory network controlling the establishment of the fly retinal mosaic has been uncovered (11) (Fig. 1B). The stochastic fate decision is initially made by R7 and is controlled by the transcription factor Spineless (Ss). Ss is stochastically turned on in 65% of R7s that then adopt the yR7 fate (10, 14). Once this decision is made cell-autonomously by R7, it is propagated to R8 in the same ommatidium so that R7 and R8 have coupled rhodopsin expression. This is achieved through induction of the pR8 fate by pR7s through Activin and bone morphogenetic protein signaling, whereas the yR8 fate is specified by default (15). R7 and R8 send their axons to the medulla, the second neuropil of the optic lobe, where they make synapses with some of the ~40,000 neurons of more than 80 different cell types that constitute the medulla (16, 17) (Fig. 1C). The medulla is retinotopically organized in ~750 columns that correspond to the ~750 ommatidia.

In contrast to the stochastic specification of photoreceptors, medulla neurons are formed following a stereotypic mode of development [reviewed in (18)]. The medulla develops from

a neuroepithelium during late third-instar larval stage and early pupation. A proneuronal wave sequentially converts single rows of neuroepithelial cells into neuroblasts, the *Drosophila* neural stem cells, until the neuroepithelium is fully consumed (19). The medulla is thus sequentially produced, similarly and concomitantly to the retina where single rows of ommatidia are sequentially added in the eye disk at the morphogenetic furrow (20–22). Once specified, medulla neuroblasts sequentially express a series of transcription factors that will command the fate of the neurons produced during each temporal window (23–25). Thus, over time, a single neuroblast is able to generate a wide repertoire of different neurons, including the entire repertoire of unicolunar neurons that are found in each medulla column with a 1:1 stoichiometry with photoreceptors (26, 27).

Connecting photoreceptors with the correct postsynaptic partners is fundamental to ensure proper color vision. Here, we investigated how the stochastic decision made by photoreceptors is propagated to the medulla to instruct the formation of yellow and pale columns in which R7 photoreceptors connect to their proper specific targets. We found that correct matching is achieved through the generation of supernumerary target neurons of each subtype. Neurons that fail to connect to their corresponding R7 photoreceptor are culled by apoptosis. Recognition of future synaptic partners is achieved using a pair of interacting cell adhesion molecules from the Dpr/DIP families expressed in R7 or their Dm8 targets (28).

We argue that competition between Dm8 subtypes for the available R7s affects both their survival and their targeting. This mechanism of elimination of supernumerary neurons upon lack of interaction between cell adhesion molecules might be a general mechanism to ensure the quantitative and qualitative matching of synaptic pairs, and to relay the stochastic decisions of sensory neurons to deeper brain regions.

Identification of subtypes of Dm8, the specific target neurons of R7

We first sought to identify the specific target neurons of the distinct R7 subtypes and thus focused on R7s' main postsynaptic partner, medulla neuron Dm8s (29, 30). The cell adhesion molecule Dpr11 is specifically expressed in yR7 during pupal development, and one of the Dpr11-binding partners, DIP γ , is expressed in a subset of Dm8s (28) (Fig. 1, D and E). Because these molecules have been proposed to play a role in establishing synaptic specificity in the optic lobe (28, 31), we reasoned that the two types of Dm8 neurons that are distinguished by DIP γ expression could correspond to the two R7 subtypes, with DIP γ ⁺ Dm8s being postsynaptic to Dpr11⁺ yR7s.

Department of Biology, New York University, New York, NY 10003, USA.

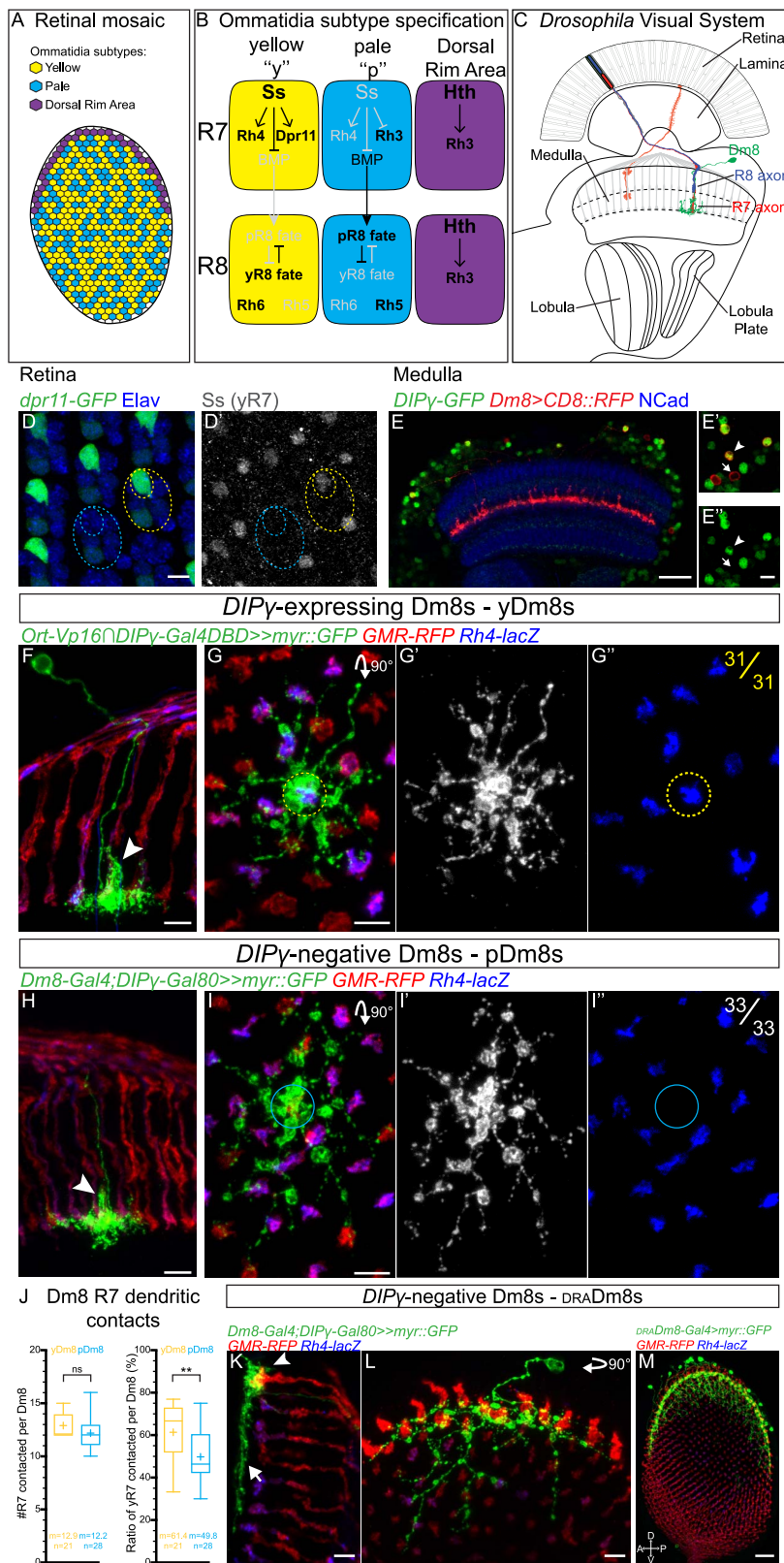
*Corresponding author. Email: cd38@nyu.edu

To test this, we first developed tools to genetically label the different populations of Dm8s according to their *DIP γ* expression. We took advantage of a *Minos* mediated integration cassette (MiMIC) inserted in the first intron of *DIP γ* (32) (fig. S1A) that faithfully recapitulates *DIP γ* expression, as confirmed by antibody stainings against *DIP γ* (fig. S1B). We swapped the green fluorescent protein (GFP) within the original MiMIC line with the Gal4 DNA binding domain to build a *DIP γ* split-Gal4 line (*DIP γ -Gal4DBD*) to label *DIP γ* -expressing Dm8s, or with Gal80 to generate a *DIP γ -Gal80* to label *DIP γ* -negative Dm8s (fig. S1A). The combination of the *DIP γ* split-Gal4 line with a hemidriver for the histamine chloride channel *ort* (*ort-C1-3-Vp16*) that is expressed in neurons postsynaptic to photoreceptors (29) labeled a large subset of Dm8 neurons (fig. S1D). To better characterize the *DIP γ* -expressing Dm8s and to look at their connectivity, we generated flip-out clones to sparsely label this Dm8 popula-

tion (Fig. 1, F and G). We confirmed that Dm8 neurons extend their dendrites in the M6 layer, where R7 projects, each contacting ~14 columns (30, 33) (Fig. 1, G and J). At the cen-

ter of their dendritic field, Dm8s extended a much more extensive dendritic branch in their “home column” along the R7 axon, from the M6 to the M4 layer that contained most of

Fig. 1. Identification of subtypes of Dm8, the specific target neurons of R7. (A) Schematic representation of the three different subtypes of ommatidia. (B) Regulatory network controlling R7 and R8 fate specification. (C) Schematic of the *Drosophila* visual system with R7 axons and their postsynaptic target Dm8 neurons. (D) *Dpr11^{M102231}* gene-trap expression in retina photoreceptors (Elav) at 25 hours APF. *Dpr11-GFP* is strongly expressed in yR7, labeled by Ss (outline in yellow circles) but absent from pR7 (cyan circles). (E) *DIP γ ^{M103222}* gene trap drives expression of GFP in the adult medulla (neuropil labeled by NCad). A subset of Dm8s (labeled by CD8::RFP) expresses *DIP γ* (arrowhead; *DIP γ* -negative Dm8s, arrow). (F) Dorsoventral view of *DIP γ* -expressing Dm8 sparsely labeled with *myr::GFP* extending a single process to the M4 layer in its home column (arrowhead). (G) Proximodistal view of a *DIP γ* -expressing Dm8 reveals that yDm8s always contact a yR7 in their home column (yellow circle, 31/31 clones). (H and I) pDm8s do not express *DIP γ* and always contact a pR7 in their home column [arrowhead in (H), cyan circle in (I); 33/33 clones]. (J) Tukey box plots representing the number of R7s contacted per yDm8s and pDm8s outside of their home column, and the percentage of these contacts being with yR7s. Mean (m) is represented by a cross. $**P = 0.0037$ (Student *t* test); ns, not significant. (K and L) A second type of *DIP γ* -negative Dm8 only contacts DRAR7s. Unlike yDm8s and pDm8s, they do not have a well-defined home column (arrowhead) and their lateral processes do not contact non-DRAR7s (arrow). (M) Proximodistal view of DRADm8s labeled by *R13E04-Gal4*. Note that DRADm8s projections are only located at the edges of the dorsal half of the medulla where DRAR7s axons are. Photoreceptors are labeled with *GMR-RFP* and yR7 with *Rh4-lacZ* in (F) to (M). Scale bars, 5 μ m [(D), (E') to (I), (K), and (L)], 20 μ m [(E) and (M)].



their synapses with R7 (29, 34) (Fig. 1F). Single-cell clonal analysis revealed that *DIPy*-expressing Dm8s always have a Rh4-expressing yR7 in their home column ($n = 31/31$ Dm8s; Fig. 1, F and G), but their lateral dendrites

contact either pR7s or yR7s (Fig. 1, G and J). Hereafter, we refer to *DIPy*-expressing Dm8s as yellow Dm8s (yDm8s).

We next characterized *DIPy*-negative Dm8s using two distinct Gal4 lines expressed in Dm8s

in combination with *DIPy-Gal80* (fig. S1F) and observed two types of neurons. One population of Dm8s was morphologically identical to yDm8s (Fig. 1H and fig. S1F). However, these neurons always had a pR7 in their home

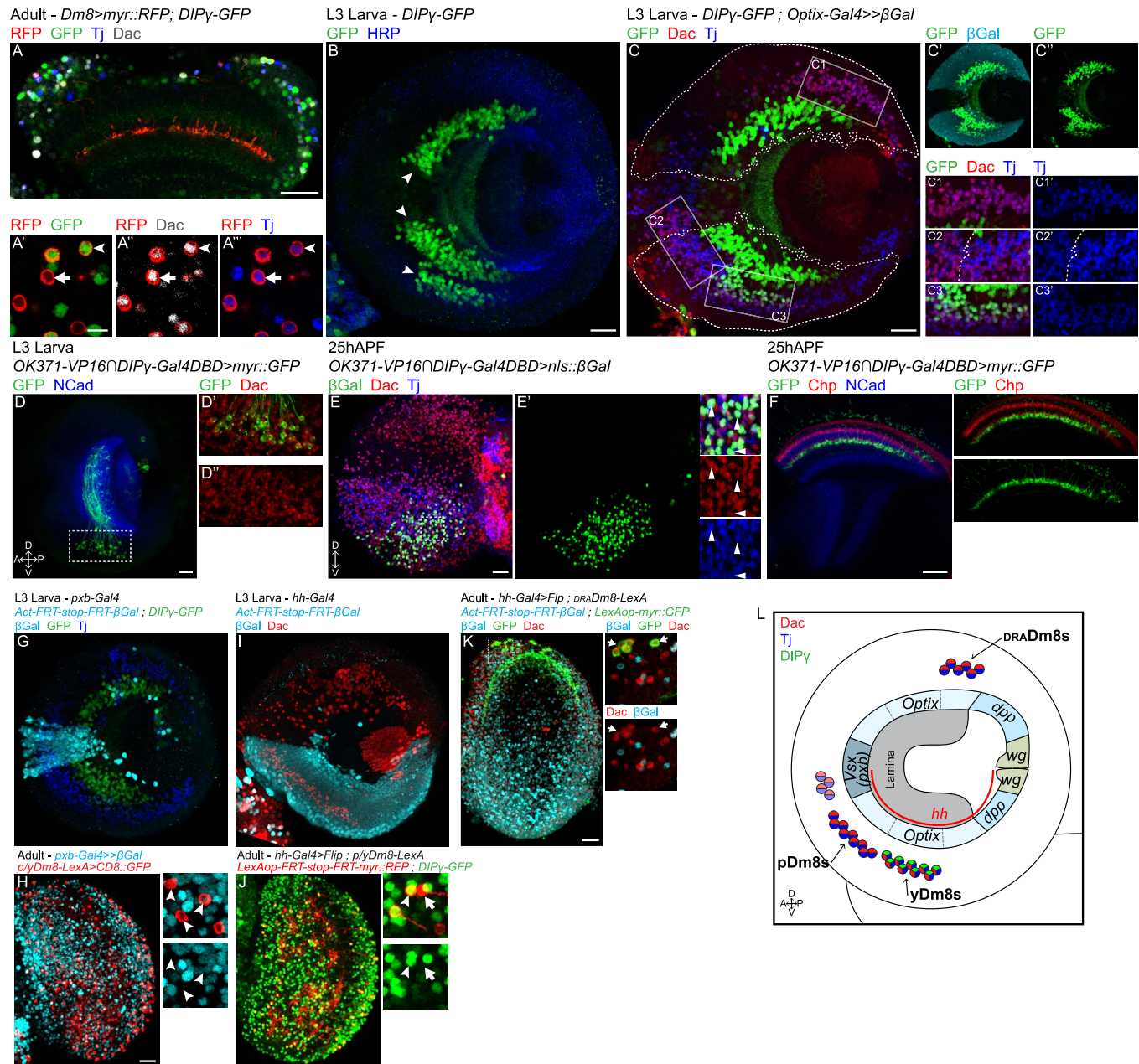


Fig. 2. Dm8 subtypes are prespecified and have distinct lineages. (A) In an adult, pDm8s and yDm8s (both labeled by *myr::RFP*, plus *DIPy-GFP* for yDm8s) express *Dac* and *Tj*. (B) *DIPy* is expressed in three clusters of neurons in late L3 optic lobe (arrowheads). *HRP* labels neuron membranes. (C) In L3, only the smaller cluster labeled by *DIPy-GFP* also expresses *Tj* and *Dac*. *Optix* lineage trace with nuclear β -galactosidase [in (C')] and outlined in (C)] revealed that this cluster is coming from the ventral part of the *Optix* domain. Three other clusters of *Tj*⁺*Dac*⁺ neurons are found in the larval optic lobe (C1 and C2). (D) The split-Gal4 line *OK371-VP16* \cap *DIPy-Gal4DBD* specifically labels the *DIPy*⁺*Dac*⁺ cell cluster in L3. (E and F) At 25 hours APF, *OK371-VP16* \cap *DIPy-Gal4DBD* is still specific to the same cluster of cells expressing *Dac* and *Tj* (arrowheads in E') that can be identified as Dm8s by morphology (F).

(G) *pxb* lineage trace in late L3 optic lobe labeled all mOPC-derived neurons with nuclear β -galactosidase. Note that none of the *DIPy-GFP* neurons are labeled by β -Gal. (H) Same lineage tool in adult in combination with a *R24F06-LexA* driving *myr::GFP* in p/yDm8s. None of the p/yDm8s express β -Gal (arrowheads). (I) *hh* lineage trace in late L3 optic lobe labeled all neurons derived from the ventral half of the OPC with nuclear β -Gal. (J) *hh* lineage tool in combination with *DIPy-GFP* labels both pDm8s (arrow) and yDm8s (arrowhead). *hh-Gal4* drives the expression of the Flip recombinase that will lead to the excision of a stop cassette within a *LexAop-RFP* reporter. Thus, only cells coming from the ventral *hh*⁺ region and expressing the p/yLexA driver will be labeled by RFP. (K) *hh* lineage trace in adult does not label *DRADm8s* (arrowhead). (L) Schematic of the distinct lineages of the three Dm8 subtypes. Scale bars, 20 μ m [(A) to (K)], 5 μ m (A').

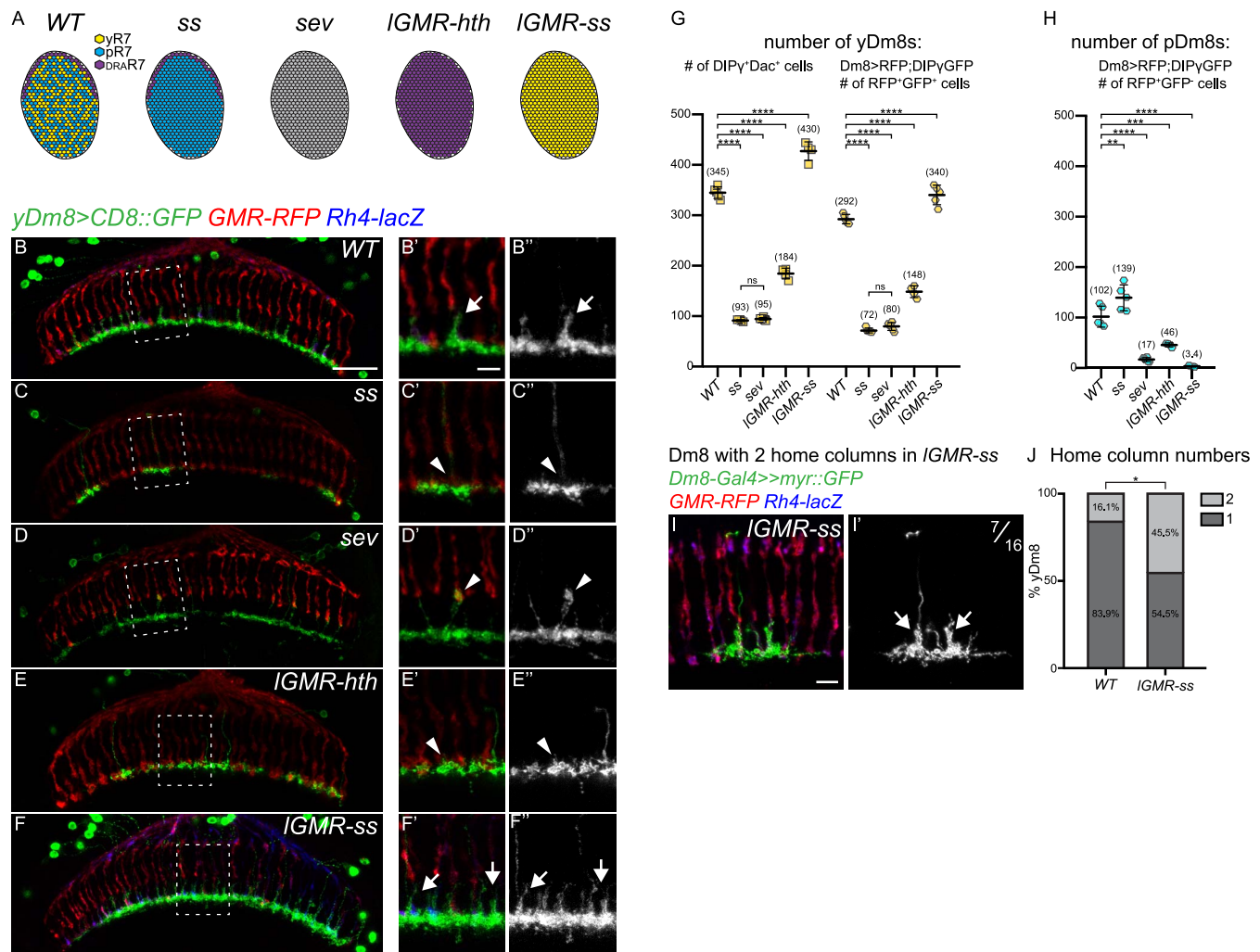


Fig. 3. Change in R7 ratio affects Dm8 subtype formation. (A) Schematic representing the R7 subtypes in the different mutant conditions. *spineless* (*ss*) mutant retinas are only composed of pale and DRA R7s; in *sevenless* (*sev*) mutants, R7 photoreceptors are not specified and thus absent. *IGMR-ss* and *IGMR-hth* are specific gain-of-function where all R7s are either all of the yellow or DRA type, respectively. (B to F) yDm8s labeled with CD8::GFP in WT (B), *ss* (C), *sev* (D), *IGMR-hth* (E), and *IGMR-ss* (F); photoreceptors with *GMR-RFP* and yR7 with *Rh4-lacZ*. Arrow and arrowheads indicate the WT or defective Dm8 process in the home column. (G and H) Quantification of the numbers of

yDm8s and pDm8s per optic lobe in *Dm8-LexA>LexAop-RFP;DIPy-GFP* animals. (G) The number of yDm8s is plotted using two different quantification: the number of DIPy⁺GFP⁺Dac⁺ cells or as the number of RFP⁺GFP⁺. (H) Quantification of the number of pDm8s (RFP⁺GFP⁺). Data are means ± SD. ***P* = 0.001, ****P* < 0.001, *****P* < 0.0001 [one-way analysis of variance (ANOVA), Tukey test]. (I) Single-cell clone of a yDm8 with two home columns (arrow) in *ss* gain-of-function (*IGMR-ss*). (J) Distribution of yDm8s with two home columns (WT, *n* = 31; *IGMR-ss*, *n* = 22). **P* = 0.0302 (Fisher exact test). Scale bars, 20 μm in B for (B) to (F), 5 μm in B' for (B') to (F') and (I).

column (*n* = 33/33 Dm8s; Fig. 1I), but, similar to yDm8s, they contacted both pR7s and yR7s outside their home column (Fig. 1, I and J). We refer to these neurons as pale Dm8s (pDm8s). Although both pDm8s and yDm8s showed a strict preference for the R7 subtype in their home column and contacted on average the same number of R7s outside their home column (Fig. 1J), the ratio of R7 subtypes contacted by their lateral dendrites was different: yDm8s connected to yR7 versus pR7 with the same frequency as the distribution of these photoreceptors (Fig. 1J, ratio of yR7 contacted = 61.4% versus yR7 frequency = 65%), whereas pDm8s had a preference for pR7s (Fig. 1J, ratio of pR7 contacted = 51.2% versus pR7

frequency = 35%). Additionally, ~15% of Dm8s from both populations harbored two main processes and thus had two home columns (fig. S1, H and I) that were always both occupied by their preferred R7 subtype.

To confirm the strict home-column pairing of yDm8s with yR7s and pDm8s with pR7s that we observed in single-cell clones, we looked at whole-mount stainings of either population (fig. S1, E and G). We never observed a yDm8 extending its main dendritic branch along a pR7 (fig. S1J, *n* = 1046) or the reverse for pDm8s (*n* = 516). We also quantified the ratio of columns occupied by a Dm8 as a home column; 88% of yR7 columns were occupied by a yDm8, whereas 96.4% of pR7

columns were occupied by a pDm8 (fig. S1J). These numbers might be a lower estimate of Dm8s' column coverage because the Gal4 lines are not fully penetrant and do not label all neurons of a given cell type [see below and (35)].

In addition to pDm8s, we also identified a second type of DIPy-negative Dm8 that only innervated DRA photoreceptors (Fig. 1, K and L). These Dm8s had a morphology distinct from that of pDm8s and yDm8s: They did not appear to have any distinctive home column; they only made tight contacts with DRAR7 termini and did not contact the M6 layer in the main part of the medulla innervated by pale and yellow R7s (Fig. 1, K and L). We also

identified a Gal4 line that specifically labeled DRADm8s, thereby confirming that these neurons were genetically different from pDm8 neurons and that DRADm8s were confined to the outer part of the dorsal half of the medulla where DRAR7 axons are located (Fig. 1M). These neurons correspond to the newly identified Dm-DRA1 neurons that were shown to be postsynaptic to DRAR7s (36).

Thus, we identified three types of Dm8s, corresponding to the three different R7 subtypes. In the main part of the medulla, most columns are occupied by the main process of a single pDm8 or yDm8 with a perfect pairing of R7 and Dm8 subtypes. Thus, the topographic organization of R7 subtypes in the retina is propagated to the medulla and mirrored by the mosaic of Dm8 subtypes.

Dm8 subtypes are prespecified and have distinct lineages

We next sought to identify the mechanisms that lead from the stochastic patterning of photoreceptors to a deterministic output in the medulla, where most of the columns are occupied by a Dm8 with a perfect matching between R7 subtypes and their respective Dm8 subtypes.

Two alternative mechanisms could allow this matching: (i) R7 subtypes could directly coordinate their fate with their Dm8 subtypes by instructing naïve Dm8s during development to adopt the appropriate fate (p versus y). This would be similar to the coordination between R7 and R8 fates, where pR7s signal to R8s within the same ommatidium and instruct them to adopt the pR8 fate (15). (ii) Alternatively, distinct Dm8 subtypes could be specified independently of R7 subtypes, such that matching would occur during later stages in development.

To distinguish between the two models, we sought to identify the origin of the distinct Dm8 subtypes and asked whether distinct subtypes form during development, before R7 innervation, or whether a single naïve type develops first. Because the Gal4 lines used to label Dm8 neurons in adult brains begin expression during late pupal development, we looked for markers expressed by adult Dm8 neurons that may also be expressed during early development. In adults, all three Dm8 subtypes expressed the transcription factors Dachshund (Dac) and Traffic jam (Tj) (Fig. 2A and fig. S2A). We first focused on identifying yDm8s during development and asked when yDm8s adopt their final subtype fate. We looked at the early expression of *DIPγ* in late L3 larval optic lobes and identified several distinct clusters of cells expressing *DIPγ* (Fig. 2B). One of these clusters also expressed Dac and Tj (Fig. 2C) and could represent the yDm8 population. The identification of larval yDm8s based on the markers expressed in

adult Dm8s assumes that their expression is maintained throughout development. We confirmed that this was indeed the case by following the $\text{Dac}^+\text{Tj}^+\text{DIP}\gamma^+$ cell cluster from L3 until we could identify these neurons as yDm8s by their morphology (Fig. 2, D to F). Thus, in late L3 stage, when medulla neurons are just born, yDm8s have already acquired their final identity and express *DIPγ*.

We then looked for other Tj and Dac double-positive cells in the developing larval optic lobe that could represent the other two Dm8 subtypes. We found four other large clusters of Dac^+Tj^+ neurons (Fig. 2C). Unlike for yDm8s, we could not trace pDm8s and DRADm8s from larva to adult because of the lack of a marker equivalent to *DIPγ*. Thus, to identify which cluster corresponded to which Dm8 subtypes, we used lineage trace experiments with the FLEXAMP memory cassette (fig. S2D) (37). When using the *tj-Gal4* line, which faithfully recapitulates Tj expression in larvae (fig. S2C), in combination with *DIPγ-Gal80*, we consistently obtained clones of pDm8s and DRADm8s (fig. S2E).

The medulla neuroepithelium is divided along the dorsoventral axis into compartments based on the expression of spatially restricted factors (26, 37): *dpp*, *Optix*, and *Vsx1* expression define the three major compartments; *Optix* is in the two arms of the medulla neuroepithelium (Fig. 2C) and *dpp* in the two lateral parts. The ventral half can also be defined by its early expression of *hedgehog* (*hh*) (Fig. 2I). We used lineage tools to identify the neuroepithelium compartments from which the different types of Dm8s originate.

We first used a lineage tool for the main medulla neuroepithelium compartment using *Optix-Gal4* (Fig. 2C) and for the central medulla neuroepithelium using *paxb-Gal4* (Fig. 2G). In larvae, the majority of Dac^+Tj^+ neurons came from the Optix region; two clusters were present in the ventral half, including the one expressing *DIPγ*. A third one was in the dorsal half (Fig. 2C), whereas a smaller cluster was in the *paxb* region (Fig. 2G). We used these lineage tools to trace yDm8s, pDm8s, and DRADm8s in adults. All three subtypes were labeled by the *Optix* lineage tool (fig. S2, F and G), whereas no Dm8s were labeled by the *paxb* lineage tool (Fig. 2H and fig. S2H). We then traced neurons coming from the ventral half of the medulla neuroepithelium by using *hh-Gal4* lineage trace (Fig. 2I) to identify whether the two Optix-derived $\text{Tj}^+\text{Dac}^+\text{DIP}\gamma^-$ clusters were the pDm8s and DRADm8s. pDm8s (and yDm8s) were labeled by the *hh* lineage trace, but none of the DRADm8s were (Fig. 2, J and K). Therefore, pDm8s come from the ventral Optix cluster, next to yDm8s, whereas DRADm8s are part of the dorsal Optix-derived cluster (Fig. 2L).

This shows that the three Dm8 subtypes come from three different neural progenitor domains and thus have distinct lineages (Fig. 2L). The distinct fates of Dm8 subtypes are thus preestablished independently of the specification of their presynaptic R7 subtype.

Change in R7 ratio affects formation of Dm8 subtypes

Because the Dm8 subtypes are specified independently of yR7 and pR7, how can the brain accommodate stochastic changes in the ratio of photoreceptor subtypes to ensure that pDm8s always connect to pR7s and yDm8s to yR7s? To address this, we used mutations that affect the specification of the different ommatidial subtypes and looked at the consequences for the formation of Dm8s (Fig. 3A). We first focused on yDm8s: In the wild type, their arborizations covered the M6 layer almost entirely, and their main branch that reached M4 in their home column could be easily identified (Fig. 3B). We examined the effect of the absence of yR7s on yDm8s by using a retina-specific allele of *spineless* in which yR7s are not specified and the main part of the retina is solely composed of pR7s (R. Johnston, personal communication). In these *ss* mutants, large areas of the M6 layer were devoid of yDm8s because of a decrease in the number of yDm8s (Fig. 3, C and G; number of yDm8s per optic lobe, WT = 345, *ss* = 93). Furthermore, remaining yDm8s lacked a home column, as seen by the absence of the typical main Dm8 arbor reaching the M4 layer (Fig. 3C). Similar effects, although varying in magnitude, could be seen in two other genetic backgrounds that lack yR7s: In *sevenless* (*sev*) mutants, where R7s are not specified (Fig. 3A), there was a similar decrease in the number of yDm8s (Fig. 3, D and G, *sev* = 95). However, in contrast with *ss* mutants, some yDm8s appeared to still have a home column, although their main processes were thinner and reached higher in the medulla to layer M3 where they wrapped around R8 termini (Fig. 3D). We also converted the entire retina into DRA ommatidia using *homothorax* gain-of-function (*IGMR-hth*, Fig. 3A) and we observed a similar, although weaker, decrease in the number of yDm8s (Fig. 3G, WT = 345, *IGMR-hth* = 184) and a total absence of yDm8s' home column (Fig. 3E).

We next performed the reverse experiment and converted the entire retina into yellow ommatidia by overexpressing Ss in photoreceptors (*IGMR-ss*, Fig. 3A). In this case, almost every R7 was occupied by a yDm8 home column (Fig. 3F) and the number of yDm8s increased by 25% (WT = 345, *IGMR-ss* = 430).

We also quantified the number of pDm8s and obtained similar results: In the absence of pR7s (*IGMR-ss*, *sev*), the number of pDm8s

dropped substantially (Fig. 3H; WT = 102, *IGMR-ss* = 3.4, *sev* = 17), whereas it increased in *ss* mutants (Fig. 3H; *ss* = 139). Taken together, these data indicate that the numbers of yDm8s and pDm8s are affected by the numbers of yR7s and pR7s in the retina.

In *ss* gain-of-function (*IGMR-ss*), most yR7s were occupied by a yDm8 home column (Fig. 3F); the 25% increase in yDm8 number alone could not account for such a strong effect (Fig. 3G). We speculated that in addition to the increase in cell number, an increase in the number of home columns covered by individual yDm8 could explain such an effect. We therefore generated single-cell flip-out clones of yDm8s in *ss* gain-of-function to look at their morphology. In the wild type, 16.1% of yDm8s had two home columns (Fig. S1, H and I), whereas in *ss* gain-of-function, 45.5% of yDm8s had two home columns (Fig. 3, I and J). We never observed yDm8s with more than two home columns. Thus, two mechanisms allow adult yDm8s to accommodate changes in the number of their R7 presynaptic partners: (i) their numbers increase or decrease overall to match the number of yR7s, and (ii) when excess yR7s are present, individual yDm8s also increase the number of

columns they occupy, such that most yR7s are covered.

Apoptosis of excess Dm8s ensures the numerical matching of R7s and Dm8s

Our data indicate that although the different types of Dm8s are prespecified, their number can be adjusted to accommodate the ratio of their presynaptic R7s. One hypothesis is that in *ss* mutants, yDm8s are produced normally, but in the absence of their yR7 partners, they are eliminated during a later stage of development, whereas when yR7s are in excess (*IGMR-ss*), all yDm8s are maintained. To test this, we first looked at the number of yDm8s early in development in *ss* mutants that lack yR7s. At 20 hours after puparium formation (APF), when the neuroblasts no longer divide and thus no more medulla neurons are produced, similar numbers of yDm8s were found in *ss* mutants and in the wild type (Fig. 4, A, C, and E; WT = 437.5, *ss* = 416.5). By 40 hours APF, however, the number of yDm8s decreased to approximately the number observed in the adult in both the wild type and *ss* mutants (Fig. 4, A to E; 40 hours APF: WT = 351, *ss* = 113; adult: WT = 345, *ss* = 93). This confirms that in the absence of yR7s, yDm8s are still

produced normally but numerical matching with yR7s happens during pupal development.

The decrease in the number of yDm8s in *ss* mutants might be due to the death of yDm8s that have failed to find their correct R7 subtype. Inhibiting apoptosis with *tj-Gal4* to misexpress the caspase inhibitor P35 in yDm8s restored the number of yDm8s in *ss* mutants to the number found at 20 hours APF (Fig. 4E; adult: WT = 345, *ss* = 93, *ss* + P35 = 406, P20: *ss* = 416.5). We could also rescue the decreased number of yDm8s in the wild type by misexpressing P35 (Fig. 4E). The final number of adult yDm8s obtained upon cell death inhibition was similar in the wild type and in *ss* mutants, and was also similar to the number in the *ss* gain-of-function and in the wild type at 20 hours APF (Figs. 3G and 4E; WT-20hAPF = 437.5, *IGMR-ss* = 430, WT + P35 = 410). This shows that during development, a fixed number of yDm8s are produced in excess but the relative number of yDm8s surviving depends on the number of their available presynaptic yR7s: Naturally occurring cell death can be rescued by providing more yR7s in *ss* gain-of-function, whereas it can be greatly increased by eliminating yR7s in *ss* or *sev* mutants. However, about 25% of

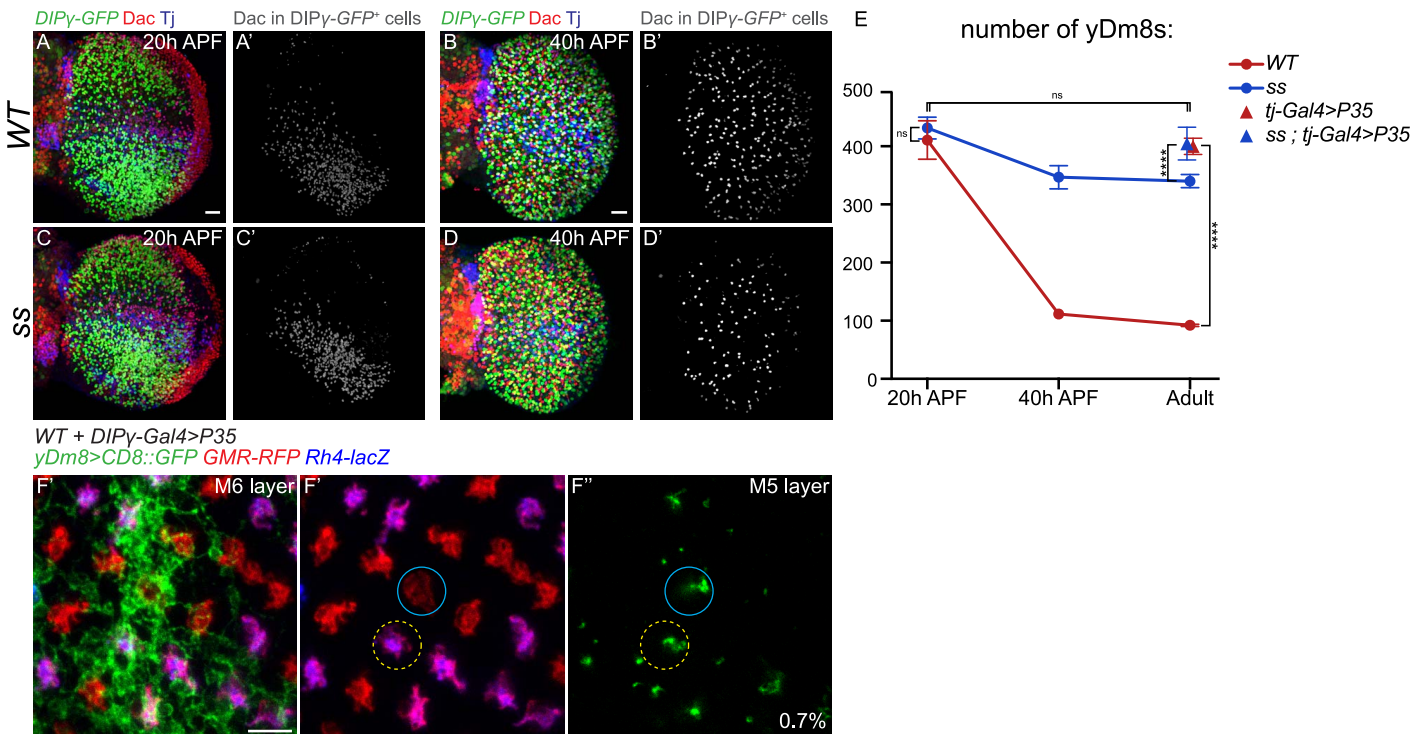


Fig. 4. Apoptosis of excess Dm8s ensures the numerical matching of R7s and Dm8s. (A to D) *DIPy-GFP* expression in the optic lobe in WT [(A) and (B)] and *ss* mutant [(C) and (D)] at 20 hours [(A) and (C)] and 40 hours APF [(B) and (D)]. yDm8s were labeled by segmenting the Dac staining from the GFP staining [(A') to (D')]. Scale bars, 20 μ m. (E) Number of yDm8s per optic lobe (*DIPy-GFP*⁺, Dac⁺ cells, *n* = 4 to 6 optic lobes per

genotype). Data are means \pm SD. *****P* < 0.0001 (one-way ANOVA, Tukey test). (F) Proximal view of yDm8s in WT upon cell death inhibition by misexpressing P35 in *DIPy*-expressing neurons. A single yDm8 mispaired with a pR7 (cyan circle) showed by dense GFP staining at the level of the M6 layer (F') and its home column at the level of the M5 layer [(F'') and (F''')] (mispaired yDm8s = 0.7%, *n* = 428). Scale bar, 5 μ m.

yDm8s are still found in the absence of any yR7s (Fig. 3G), which suggests that some yDm8s that are not connected to yR7s can still survive. This could be because they received trophic support from other neuronal partners, either from neurons that are normally connected to R7s (e.g., their postsynaptic target Tm5b) or through ectopic connections that they form with other neurons (e.g., pR7).

We propose that this mechanism is sufficient to obtain the perfect matching observed in the wild type, and that cell death plays an essential role in coordinating the size of the Dm8 populations with the ratio of y/p R7 subtypes in the retina.

Physiological apoptosis regulates Dm8 wiring

We next tested whether the physiological cell death might be important for the proper wiring of yDm8s by looking for mispairing of yDm8s with pR7s when cell death was abolished in an otherwise wild-type background. We did observe a very low but significant frequency of yDm8s mispaired with pR7s (Fig. 4F and fig. S1E; WT + *P35* = 0.7%; $n = 428$). This suggests that the great majority of undead yDm8s still manage to integrate into the proper circuitry.

yDm8 morphology and survival are affected in *DIPγ* and *dpr11* mutants

We next sought to identify the mechanisms that control the pairing of Dm8s with their specific R7 subtypes and to investigate the role of *Dpr11* and *DIPγ* in the process. Dprs and DIPs are two closely related families of immunoglobulin-containing cell adhesion molecules (38). Each of the 21 Dprs binds to one or several of the 11 DIPs, and these interactions are required for their neurogenic function (39, 40). Because of the striking complementary expression pattern of Dpr and DIP pairs in synaptic partners, these families have been proposed to play a role in synaptic partner matching (28, 31, 39). *Dpr11* and *DIPγ* are ideal candidates for the matching of yR7s and yDm8s: *dpr11* expression is specific to yR7s (28) (Fig. 1D) and depends on *ss*, as *dpr11* is lost from yR7s in *ss* mutant retinas (Fig. 5A), whereas *ss* overexpression in photoreceptors is sufficient to induce *dpr11* expression in all R7s (Fig. 5B). *dpr11* is widely expressed in the optic lobe, especially in adult brains (fig. S3A). However, at 25 hours APF, around the time *dpr11* expression peaks in yR7s, it is relatively restricted to yR7s in the M6 layer while still being broadly expressed in other medulla layers (Fig. 5C). By that time, yDm8s have already reached the M6 layer and have contacted R7s (Fig. 5C) but do not have an obvious phenotype in *DIPγ* mutants (Fig. 5D).

In adults, however, the mutant phenotypes for *dpr11* or *DIPγ* were quite obvious: yDm8s' number and innervation of the M6 layer were

decreased (Fig. 5, E and J; WT = 345, *DIPγ* = 130, *dpr11* = 132). This reduction was also due to apoptosis during development, as the *DIPγ* phenotype could be rescued by misexpressing *P35* in yDm8s (Fig. 5J; *DIPγ* + *P35* = 432). This phenotype is similar to what was reported for mutants for *DIPα* and its two Dpr partners, *dpr6* and *dpr10*, where a proportion of the three *DIPα*-expressing Dm neuron types were shown to undergo increased apoptosis (39).

Additionally, yDm8 morphology was affected in both mutants. *DIPγ* mutant yDm8s failed to extend a proper process in their home column (Figs. 3B and 5E) and only had a short protrusion at the center of their dendritic field (Fig. 5, G and H). In *dpr11* mutants, yDm8s had a similar but weaker phenotype and extended a thin process in their home column (Fig. 5, G and I). Rescuing cell death in yDm8s was not sufficient to rescue the morphology of their dendritic extension in their home column (fig. S3B).

Thus, *dpr11* and *DIPγ* mutants phenocopy the loss of yR7s (Fig. 3), supporting a model that yDm8s in these mutants are unable to recognize yR7s and thus do not receive the trophic support required for their survival. If this was indeed the case, (i) overexpressing *dpr11* in R7s should compensate for the loss of yR7s in *ss* mutants, and (ii) the loss of *DIPγ* should be epistatic over the *ss* gain-of-function. Overexpression of *dpr11* in photoreceptors increased the number of yDm8s in the wild type and could also rescue yDm8 cell death and morphology in *ss* mutants [Fig. 5K (*IGMR-dpr11* = 400, *ss* + *IGMR-dpr11* = 382) and fig. S3D]. Conversely, increasing the number of yR7s using *ss* gain-of-function was not sufficient to rescue cell death of yDm8s in a *DIPγ* mutant background [Fig. 5K (*DIPγ* = 130, *DIPγ* + *IGMR-ss* = 154) and fig. S3C].

DIPγ and *Dpr11* regulate the pairing between yR7s and yDm8s

These results imply that *Dpr11* and *DIPγ* mediate the strict matching of yDm8s with yR7s. If true, we would expect to observe mispairing of the remaining yDm8s with pR7s in either mutant because of the inability of yDm8 to recognize yR7s. In the wild type, yDm8 home columns were always located along yR7 and we never observed mispairing with a pR7, either in single-cell clones (Fig. 1, E and F) or whole mounts (Fig. 6A; number of yDm8s, $n = 1046$). We first tested whether *DIPγ* and *dpr11* mutants exhibited defects in yDm8 and yR7 pairing. We did observe mispairing in both mutants (Fig. 6, B and C) where ~5% of surviving yDm8s were paired with pR7s (Fig. 6D; *DIPγ* = 4.7%; *dpr11* = 4.8%). The ratio of yR7s contacted by *DIPγ* mutant yDm8 lateral dendrites was also decreased (fig. S4B; WT = 61.4%, *DIPγ* = 51.1%) without affecting their

overall dendritic field size (fig. S4A; WT = 12.9, *DIPγ* = 12.5).

We next tested whether *DIPγ* overexpression in pDm8s or *dpr11* in pR7s was sufficient to generate mispairing of Dm8s with R7s. When *DIPγ* was sparsely overexpressed in pDm8 MARCM clones from late third larval instar stage onward using *tj-Gal4*, pDm8s were always mispaired with yR7s (Fig. 6E, $n = 12/12$).

We also misexpressed *DIPγ* in two other medulla neuron types: Dm11 and Dm12. Dm11s project to the M6 layer and have multiple processes going along multiple R7s (fig. S2E) but do not show a preference for a R7 subtype (fig. S4C). *DIPγ* overexpression in Dm11s was sufficient to induce these processes to exclusively occupy yR7 columns (fig. S4D). However, misexpression of *DIPγ* had no effect on Dm12 neurons (fig. S4, E and F).

We also overexpressed *dpr11* in all photoreceptors and looked at whether this was sufficient to create mispairing between yDm8s with pR7s. Indeed, 20% of yDm8s extended their home column in pR7 columns (Fig. 6F, $n = 5/23$). Taken together, these data show that *Dpr11*-*DIPγ* interaction is sufficient to promote pairing of Dm8s with yR7s, whereas lack of *Dpr11* and *DIPγ* only causes 5% of yDm8s to mispair with pR7s.

Two hypotheses could explain the discrepancy between the requirement and the sufficiency of *dpr11* and *DIPγ*: (i) In the absence of *DIPγ*, pDm8s are unaffected and thus target pR7s, leaving no space for mutant yDm8s to target pR7s. One could test this by looking at yDm8 pairing in a *DIPγ* mutant where pDm8s were ablated. Because this experiment was not technically feasible, instead of removing the competition with pDm8s, we allowed *DIPγ* mutant yDm8s to compete with both pDm8s and wild-type yDm8s. We looked at the pairing of *DIPγ* homozygous mutant yDm8s in a mosaic animal using MARCM where most yDm8s were heterozygous for *DIPγ*. Because *DIPγ* mutant yDm8s would normally undergo apoptosis (see below and fig. S6), we rescued cell death by misexpressing *P35* in the mutant clones. In these conditions, half of the *DIPγ* mutant yDm8s mispaired with pR7s (Fig. 6H, $n = 7/14$), whereas in single-cell flip-out clones in whole *DIPγ* mutants, we did not observe a single mispaired yDm8 (Fig. 6G; $n = 0/15$). Thus, in this experimental setup where pR7 and yR7 columns are equally accessible, yDm8s evenly distribute between the two, which suggests that *DIPγ* is required for yDm8s to pair with the proper R7 subtype. (ii) A second explanation could be that because two-thirds of yDm8s die during early pupal development in *DIPγ* mutants after they are unable to find a yR7, they would be able to mistarget if cell death was rescued. Indeed, when we rescued cell death in *DIPγ* mutants, there was a factor of 3 increase in yDm8 mispairing (Fig. 6D;

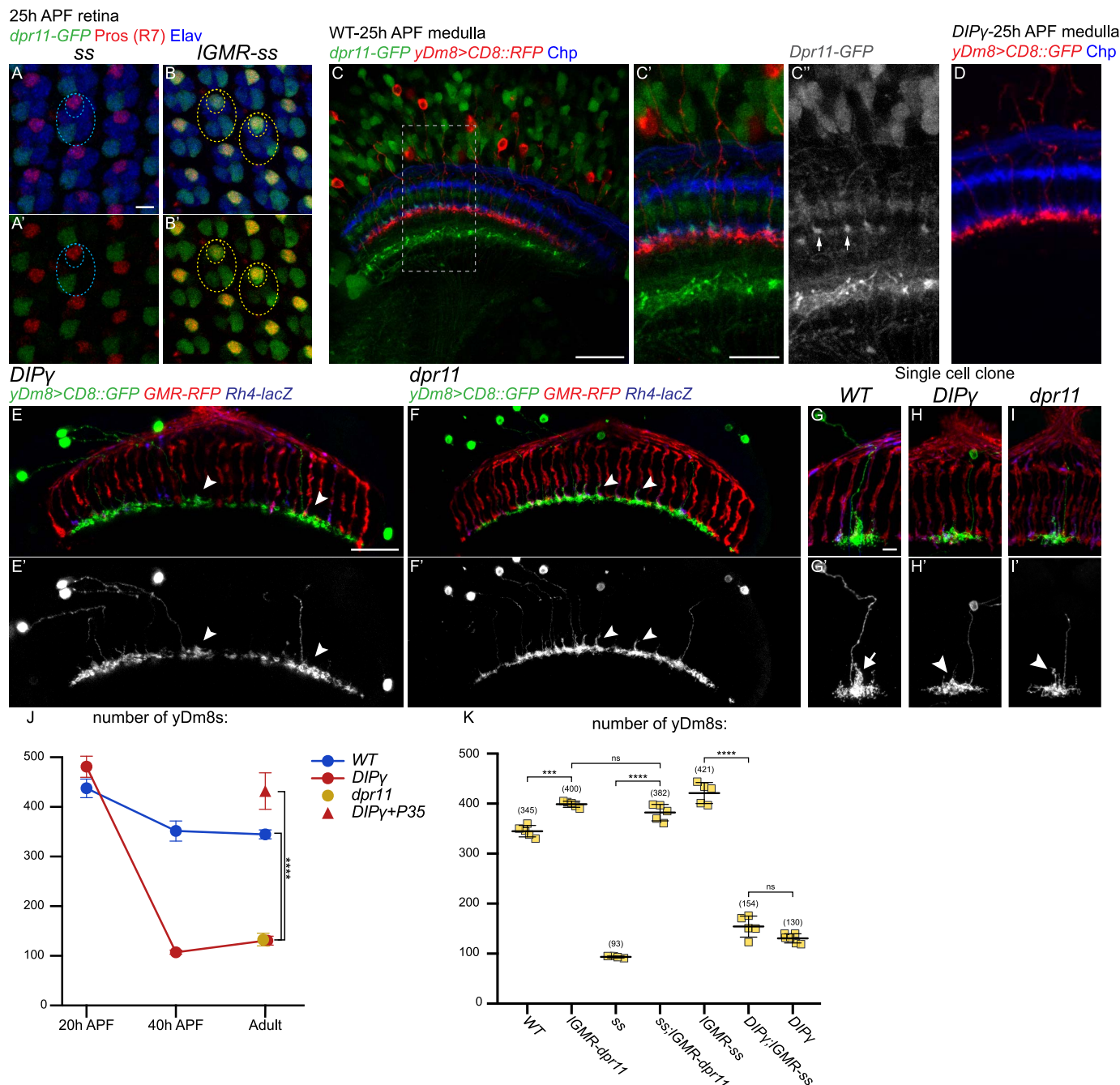


Fig. 5. yDm8 morphology and survival are affected in DIP γ and *dpr11* mutants. (A and B) *dpr11-GFP* expression in retinas at 25 hours APF in ss mutant (A) or ss gain-of-function (B). Prospero labels R7s; Elav labels photoreceptors. Large ellipse indicates a single ommatidium and smaller ellipse denotes the R7 cell. Note that *dpr11-GFP* expression is not lost in a single outer photoreceptor in ss mutant, whereas weak *dpr11-GFP* is also seen in a few outer photoreceptors in the ss gain-of-function. (C) *dpr11-GFP* expression in medulla at 25 hours APF. The split-Gal4 line *DIPy-Gal4DBD* \cap *OK371-Vp16* drives *CD8::RFP* in yDm8s. In the M6 layer, *dpr11-GFP* is mainly seen in yR7s

axons (arrows). (D) DIPy mutant yDm8s labeled with *CD8::GFP* at 25 hours APF. (E and F) Dorsal/ventral view of yDm8s in DIPy (E) and *dpr11* (F) mutants. Arrowheads indicate morphological defects in yDm8 home columns. (G to I) Sparsely labeled yDm8s in WT (G), DIPy mutant (H), and *dpr11* mutant (I); arrow and arrowheads indicate the WT or defective Dm8 process in the home column. (J and K) Number of yDm8s per optic lobe (DIPy-GFP⁺, *Dac*⁺ cells, $n = 4$ to 7 optic lobes per genotype). Data are means \pm SD. *** $P < 0.001$, **** $P < 0.0001$ (one-way ANOVA, Tukey test). Scale bars, 5 μ m [(A), (B), (C'), (D), and (G) to (I)], 20 μ m [(C), (E), and (F)].

DIPy = 4.7%, DIPy + P35 = 14.7%). However, because not every yDm8 that was prevented from dying mispaired with a pR7, this suggests that cell death is not the result of the culling of mispaired yDm8s; rather, the implication is

that preventing cell death makes them more competent to compete with pDm8s to occupy pR7 columns.

Taken together, our data indicate that *Dpr11* and DIPy mediate the pairing of yDm8s with

their presynaptic partner yR7s. It is noteworthy that we did not find any other DIPs expressed in pDm8s or pR7s, whereas Dm8s express multiple Dprs based on RNA sequencing data (17). This suggests that the matching between pR7

and pDm8s uses cell adhesion molecules distinct from Dprs and DIPs.

Targeting of the M6 layer by yDm8s is not affected in $DIP\gamma$ mutants and in DIP gain-of-function

On the basis of their layer-specific expression in the medulla and misexpression experiments, it was proposed that Dprs and DIPs regulate layer targeting (39). In either *dpr11* or $DIP\gamma$ mutants, yR7s or yDm8s do not mistarget but instead elaborate processes in the appropriate M6 layer (Fig. 5, E and F). Because the mistargeted cells could have been those eliminated by apoptosis, we looked at yDm8s in $DIP\gamma$ mutants where cell death was abolished, as well as during development when apoptosis happens. We did not observe mistargeting to another layer in either case (Fig. 5D and fig. S3B). To test whether *DIP* overexpression was sufficient to mistarget neurons to an improper layer, we misexpressed *DIPs* in different neuronal populations. As described above, overexpression of $DIP\gamma$ in Dm12s that normally innervate the M3 layer was not sufficient to make them target the M6 layer or other layers where $DIP\gamma$ is expressed (fig. S4, E and F). We also tested whether replacing $DIP\gamma$ by other DIPs normally expressed in different layers would be sufficient to retarget yDm8s to these

layers (fig. S5). Overexpression of $DIP\delta$ in $DIP\gamma$ mutant yDm8s had no effect (fig. S5E), but overexpression of $DIP\alpha$ led some yDm8s to send small processes to the M3 and M7 layers (fig. S5C), where $DIP\alpha$ and its two ligands are expressed (fig. S5, B, F, and G). Thus, the targeting of yDm8s to the M6 layer is independent of Dpr11- $DIP\gamma$ interaction, and yDm8s cannot be efficiently retargeted to different layers by ectopic expression of other DIPs. However, the extension of their home column requires $DIP\gamma$, and ectopic expression of $DIP\alpha$ leads to the formation of small extensions reminiscent of Dm8s' home column. Taken together, these results suggest that DIPs do not play an important role in layer targeting in the visual system but instead are involved in the matching of synaptic pairs.

Competitive interactions between Dm8s regulate survival and wiring

The mispairing of yDm8s was much enhanced when they had to compete for targeting with wild-type yDm8s (see above). We thus asked whether survival was also affected by competition among Dm8s. We generated $DIP\gamma$ yDm8 mutant MARCM clones in either a $DIP\gamma$ heterozygous background or in a $DIP\gamma$ mutant background where yDm8s both outside and within the clone had identical genotypes (fig.

S6A). If competition among yDm8s played a role in survival, we would expect to see differences in clone size. In the heterozygous background, we obtained small clones of yDm8s (fig. S6B; yDm8s per clone: 2 ± 2.6 , mean \pm SD) whereas in $DIP\gamma$ mutant background, we obtained clones of significantly larger size (fig. S6B; yDm8/clone: 10.8 ± 3.5). Because these experiments rely on generating clones of the same size, we controlled for clone size by quantifying the number of pDm8s per clone. In both conditions, we obtained similar numbers of pDm8s (fig. S6B; pDm8s per clone: WT background, 29.6 ± 10.3 ; $DIP\gamma$ background, 28.6 ± 6.3), thereby confirming that the difference in yDm8 number comes from a difference in survival. Thus, $DIP\gamma$ mutant yDm8s are much more likely to survive when they compete with $DIP\gamma$ mutant yDm8s rather than with wild-type yDm8s.

We noticed that affecting yDm8s sometimes yielded unexpected effects on pDm8s that could also be explained by competitive interactions among Dm8s: In $DIP\gamma$ and *dpr11* mutants, the number of pDm8s increased (fig. S6D; WT = 102, $DIP\gamma$ = 135, *dpr11* = 136). Because neither $DIP\gamma$ nor *dpr11* is expressed in pDm8s, the increase in pDm8 number might result from the decrease in the number of yDm8s. We thus explored the nonautonomous

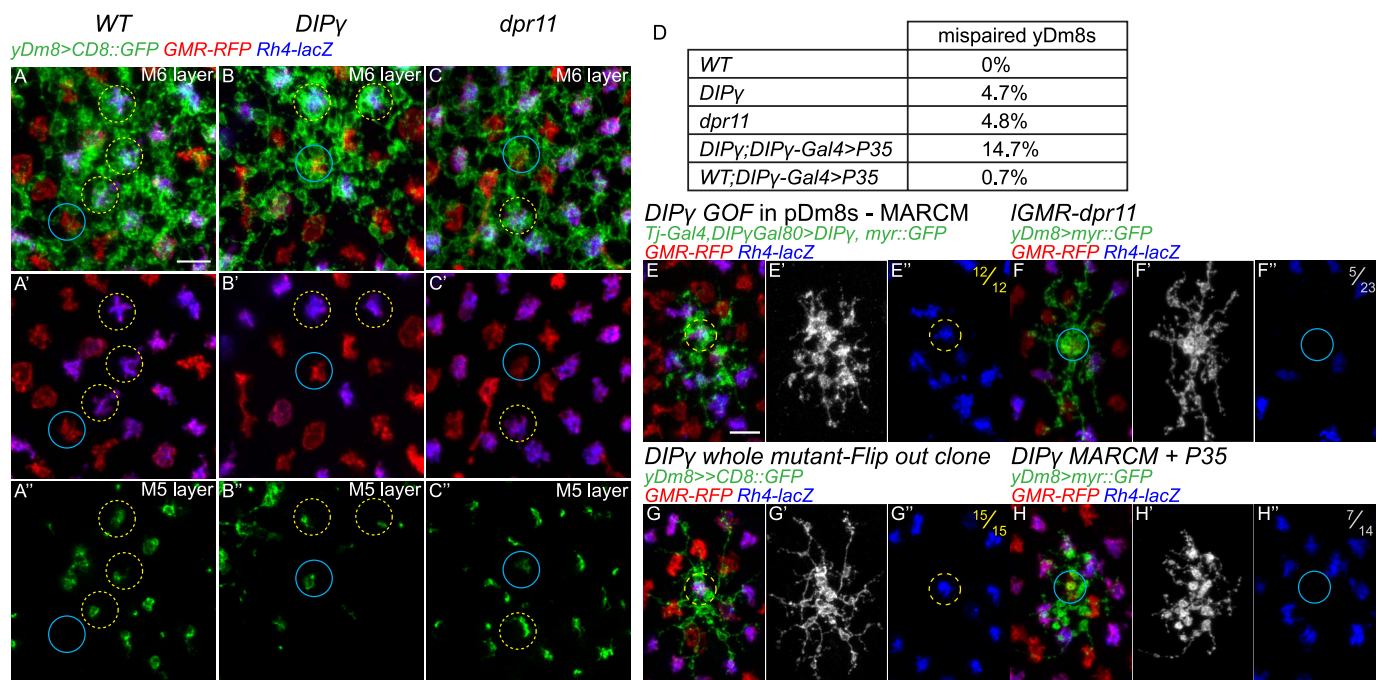


Fig. 6. $DIP\gamma$ /Dpr11 regulates the pairing between yR7s and yDm8s. (A to C) Proximal view of yDm8s labeled with CD8::GFP in WT (A), $DIP\gamma$ mutant (B), and *dpr11* mutant (C), either at the level of the M6 layer [(A) to (C) and (A') to (C')] or M5 layer [(A'') to (C'')]. Some yR7 or pR7 columns are indicated by yellow and cyan circles, respectively. In WT, yDm8s occupy most yR7 columns but never occupy pR7 columns (A''). In $DIP\gamma$ and *dpr11* mutants, many yR7 columns are devoid of yDm8s, and some yDm8s contact pR7s (cyan circle). (D) Quantification of yDm8s mispairing with pR7s (number of yDm8s: WT, $n = 1046$; $DIP\gamma$, $n = 223$;

dpr11, $n = 478$; $DIP\gamma$; $DIP\gamma$ -Gal4>P35, $n = 251$; WT; $DIP\gamma$ -Gal4>P35, $n = 428$). (E) $DIP\gamma$ overexpression in pDm8 MARCM clones is sufficient for mispairing of pDm8s with yR7s ($n = 12/12$). (F) yDm8 flip-out clone with *dpr11* overexpressed in all photoreceptors. Some yDm8s are mispaired with pR7s ($n = 5/23$). (G) yDm8 flip-out clone in $DIP\gamma$ background; 100% of the yDm8 clones obtained are paired with yR7s ($n = 15/15$). (H) $DIP\gamma$ mutant MARCM clone of a yDm8 expressing myr::GFP and P35 in an otherwise heterozygous background; 50% of the yDm8 clones obtained are mispaired with pR7s ($n = 7/14$). All scale bars, 5 μ m.

effects on pDm8 survival. We first asked whether promoting survival of yDm8s would affect pDm8 survival when we increased the number of yDm8s without affecting pR7 specification: In *dpr11* gain-of-function (*IGMR-dpr11*), the number of yDm8s increased, with a corresponding >50% decrease in the number of pDm8s (fig. S6D; WT = 102, *IGMR-dpr11* = 41). We obtained the same effect in the absence of any yR7s when *dpr11* was overexpressed in a *ss* mutant (fig. S6D; *ss* = 139, *ss+IGMR-dpr11* = 32). We also asked whether decreasing the number of yDm8s (by mutating *DIPγ*) in a *ss* gain-of-function would have an effect on the survival of pDm8s (fig. S6D), almost all of which normally die (Fig. 3G). The decreased number of yDm8s was accompanied by an increase of pDm8s (fig. S6D; *IGMR-ss* = 3.4, *DIPγ + IGMR-ss* = 29). Taken together, these results show that the size of the pDm8 population is affected by the number of yDm8s. Dm8s must compete for targeting, and affecting the size of one population affects the survival of the other. This competition might be regulated by self-avoidance mechanisms (41) and would explain why there is never more than one Dm8 home column per R7.

Discussion

In *Drosophila*, patterning of the mosaic of ommatidial subtypes is established by sequential steps and is initiated in R7 photoreceptors by a single transcription factor, Ss. This initial decision is then transmitted to the R8 of the same ommatidium so that R7 and R8 have paired rhodopsin expression (15). In contrast, our results show that the mechanism responsible for the coordination of R7 subtype specification with their main postsynaptic target in the brain is different. Dm8 subtype specification does not depend on direct induction from R7s, and each Dm8 subtype is produced independently from their corresponding R7s. We found that the matching occurs during a later stage of circuit formation, after the specification of the different components of the circuit. After matching with their R7 subtypes, controlled by *Dpr11/DIPγ* for yR7s/yDm8s, supernumerary Dm8s are culled by apoptosis. Because the ratio of ommatidial subtypes can vary among individuals (42), this mechanism allows the perfect matching that is observed in adults, where most R7s, if not all, are innervated by a single Dm8 of the proper type. This developmental plasticity provides a powerful mechanism to transmit the stochastic formation of the photoreceptor mosaic to the deterministic patterning of the brain.

Role of *DIPγ* and *Dpr11* in synaptic partner pairing

The complementary expression of Dpr and DIP binding pairs in synaptic partners raised

the possibility that these proteins might be the long-sought “Sperry molecules” that act as molecular tags to instruct synaptic specificity (28, 31, 43). Here, we provide evidence that *DIPγ* in yDm8 and *Dpr11* in yR7 instruct synaptic partner matching. Similar to the genetic removal of yR7s, loss of either *DIPγ* or *Dpr11* leads to apoptosis of yDm8s; this suggests that in their absence, yDm8s are unable to connect to yR7s and to receive from yR7s the trophic support required for their survival. In both mutants, this is accompanied by limited mispairing of yDm8s with pR7s, whereas ectopic expression of these molecules is sufficient to create mispairing between R7 and Dm8 subtypes. However, targeting to the proper M6 layer of the medulla is not affected in these mutants, nor is the dendritic size of yDm8s. Thus, we propose that Dprs and DIPs act during a later step of circuit formation to allow distinct neurons that project to the same layer to recognize their appropriate synaptic partners.

Analysis of other Dpr/DIP pairs in and outside the visual system supports this view: In the medulla, loss of either *DIPα* or its two ligands, *Dpr6* and *Dpr10*, results in apoptosis of a proportion of three Dm neuron types that express *DIPα*, likely because these neurons are unable to recognize their targets (39). In the olfactory system, loss of DIPs leads to the disorganization of olfactory glomeruli (44). At the larval and adult neuromuscular junction, *DIPα* is expressed in a subset of motoneurons, whereas its binding partner *Dpr10* is in muscles (45, 46). Loss of either leads to the partial loss of the innervation of the muscle by motoneurons. During development, mutant adult motoneurons extend normal filopodia that target the proper muscles. However, these filopodia fail to be maintained, likely because they are unable to recognize the proper muscles in the absence of these molecules (45). In the lamina, loss or misexpression of DIPs leads to ectopic synapse formation (47).

Taken together, this work suggests that Dprs and DIPs establish synaptic specificity. The difference in phenotypes in different systems (e.g., survival; mistargeting or loss of axonal branches) might reveal the different requirements for such molecules in distinct circuits.

Apoptosis as a mechanism for numerical matching of neuronal pairs

Programmed cell death facilitates quantitative matching of synaptic partners (48). For example, target-derived nerve growth factor (NGF) promotes survival of sympathetic and sensory neurons (49, 50). This discovery led to the development of the neurotrophic theory, which states that neurons are produced in excess and that competition for limited trophic support allows for the numerical matching of afferents with their targets in the periphery. Here, we

provide evidence that a similar phenomenon happens in the *Drosophila* central nervous system. We show that yDm8s are produced in excess, and that around 25% are eliminated by apoptosis during normal development. This cell death could be fully rescued by increasing the number of yDm8 afferents, or conversely be aggravated by decreasing the number of yR7s. Because *DIPγ* and *dpr11* mutants phenocopy the loss of yR7, uncovering a link between synaptic pairs matching and survival, this argues that the numerical matching of pairs of R7 and Dm8 is obtained by apoptosis of unmatched Dm8s.

In other parts of the visual system and of the brain, neuronal survival is not affected in DIP mutants (44–47). Thus, cell death might reveal circuit-specific properties of the formation of the visual system. For yDm8s, the dependency on targeting for survival allows the 1:1 matching of yDm8s and yR7s. For other multicolumnar neurons, one can only speculate about the function of normally occurring cell death. The medulla is composed of ~750 retinotopically organized columns that can be considered as repetitive microcircuits that each compute information from a single point of the visual field. Each column needs to be innervated by the proper neurons in the proper amount. Thus, creating more multicolumnar neurons that require target-derived trophic support to survive might allow the complete innervation of the medulla while having the optimal number of neurons.

Materials and methods

Drosophila strains

Flies were kept on standard cornmeal medium at 25°C 12-hour light/dark cycles (except when otherwise specified).

Dpr11^{M102231}(#40181), *DIPγ*^{M103222}(#35928), *GMR24F06-Gal4*(#49087), *GMR24F06-LexA*(#52695), *GMR13E04-Gal4*(#48565), *GMR13E04-LexA*(#52457), *OK371-Gal4*(#26160), *tj-Gal4*^{NP1624}(K#104055), *Optix-Gal4*^{NP2631}(K#104266), *hh-Gal4*(#67046), *pxb*^{M105058}(#37891), *DIPα::GFP*^{M102031}(#60523), *DIPδ::GFP*^{M108287}(#60558), *Dpr6::GFP*^{M101358}(#59287), *Dpr10::GFP*^{M103557}(#59807), *Dpr12::GFP*^{M101695}(#60171).

R57C10-FLPG5::PEST(#64089), *10XUAS-FRT-stop-FRT-myr::smGdP-FLAG*(#62130), *10XUAS-FRT-stop-FRT-myr::smGdP-HA*, *10XUAS-FRT-stop-FRT-myr::smGdP-V5-THS-10XUAS-FRT-stop-FRT-myr::smGdP-FLAG*(#64085), *GMR-myr::RFP*(#7120, #7122, #7123), *Rh4-lacZ*(#8480, #8481), *UAS-nls::βGal*(#3955), *LexAop-FRT-stop-FRT-CD8::GFP*(#57588), *UAS-P35*(#6298, #5072), *10XUAS-myr::GFP*(#32198), *10X UAS-CD8::RFP*(#32219), *Act-FRT-stop-FRT-nls::βGal*(#6355), *13xLexAop2-CD8::GFP*(#32205, #32203), *13xLexAop2-myr::GFP*(#32209), *13xLexAop2-6xmCherry*(#52272), *R11C05-LexA*(#54608), *R47G08-Gal4*(#50328), *FRT82b*, *Tub-Gal80*(#5135), *Tub-Gal80ts*(#7108), *FRT19A*

(#1709), *hs-Flp*, *FRT19A*, *Tub-Gal80*(#5132), *20XUAS-FLPG5::PEST*(#55807), *13aLexAop2-FRT-stop-FRT-myr::smGdP-V5*(#62107), *Df(3R)Exel7330*(#7985), *sev¹⁴*(#10546), *Act-FRT-stop-FRT-LexA*.

OK371-Vp16, *Otc2b-Gal4*, *Otc1-3-Vp16* from Chi-Hon Lee, *ss^{ΔR7}* from R. Johnston, *dpr11^{null}* and *DIPγ^{null}* from L. Zipursky (39), *DIPγ-Gal4^{M103222}* from H. Bellen (28).

From this study: *IGMR-ss*, *IGMR-hth*, *IGMR-dpr11*, *LexAop2-FRT-stop-FRT-myr::RFP*, *pxb-T2A.Gal4^{M105058}*, *10XUAS-DIPα*, *10XUAS-DIPδ*, *10XUAS-DIPγ*, *DIPγ-Gal4DBD* and *DIPγ-Gal80*.

The detailed genotype for each figure is given in table S1.

Immunohistochemistry

Fly brains were dissected in ice-cold PBS and fixed for 3 hours in 4% formaldehyde (v/w) in 1× PBS at 4°C. After a 30-min wash in PBST (1× PBS + 0.4% Triton X-100 and 0.5% goat serum), brains were incubated for 2 days in primary antibodies diluted in PBST, followed by 2 days with secondary antibodies diluted in PBST. After washes, brains were mounted in Slowfade and imaged on either a Leica SP5 or SP8 confocal. The antibodies used are listed in table S2.

REFERENCES AND NOTES

- R. J. Johnston Jr., C. Desplan, Stochastic mechanisms of cell fate specification that yield random or robust outcomes. *Annu. Rev. Cell Dev. Biol.* **26**, 689–719 (2010). doi: [10.1146/annurev-cellbio.100109-104113](https://doi.org/10.1146/annurev-cellbio.100109-104113); pmid: 20590453
- L. Buck, R. Axel, A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **65**, 175–187 (1991). doi: [10.1016/0092-8674\(91\)90418-X](https://doi.org/10.1016/0092-8674(91)90418-X); pmid: 1840504
- J. Nathans, D. Thomas, D. S. Hogness, Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science* **232**, 193–202 (1986). doi: [10.1126/science.2937147](https://doi.org/10.1126/science.2937147); pmid: 2937147
- A. Roorda, D. R. Williams, The arrangement of the three cone classes in the living human eye. *Nature* **397**, 520–522 (1999). doi: [10.1038/17383](https://doi.org/10.1038/17383); pmid: 10028967
- K. J. Ressler, S. L. Sullivan, L. B. Buck, A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* **73**, 597–609 (1993). doi: [10.1016/0092-8674\(93\)90145-G](https://doi.org/10.1016/0092-8674(93)90145-G); pmid: 7683976
- P. Mombaerts et al., Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996). doi: [10.1016/S0092-8674\(00\)81387-2](https://doi.org/10.1016/S0092-8674(00)81387-2); pmid: 8929536
- P. Mombaerts, Axonal wiring in the mouse olfactory system. *Annu. Rev. Cell Dev. Biol.* **22**, 713–737 (2006). doi: [10.1146/annurev-cellbio.21.012804.093915](https://doi.org/10.1146/annurev-cellbio.21.012804.093915); pmid: 17029582
- K. J. Ressler, S. L. Sullivan, L. B. Buck, Information coding in the olfactory system: Evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245–1255 (1994). doi: [10.1016/0092-8674\(94\)90015-9](https://doi.org/10.1016/0092-8674(94)90015-9); pmid: 7528109
- N. Franceschini, K. Kirschfeld, B. Minke, Fluorescence of photoreceptor cells observed in vivo. *Science* **213**, 1264–1267 (1981). doi: [10.1126/science.7268434](https://doi.org/10.1126/science.7268434); pmid: 7268434
- M. F. Wernet et al., Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**, 174–180 (2006). doi: [10.1038/nature04615](https://doi.org/10.1038/nature04615); pmid: 16525464
- J. Rister, C. Desplan, D. Vasiliauskas, Establishing and maintaining gene expression patterns: Insights from sensory receptor patterning. *Development* **140**, 493–503 (2013). doi: [10.1242/dev.079095](https://doi.org/10.1242/dev.079095); pmid: 23293281
- M. F. Wernet et al., Homothorax switches function of Drosophila photoreceptors from color to polarized light sensors. *Cell* **115**, 267–279 (2003). doi: [10.1016/S0092-8674\(03\)00848-1](https://doi.org/10.1016/S0092-8674(03)00848-1); pmid: 14636555
- M. F. Wernet et al., Genetic dissection reveals two separate retinal substrates for polarization vision in Drosophila. *Curr. Biol.* **22**, 12–20 (2012). doi: [10.1016/j.cub.2011.11.028](https://doi.org/10.1016/j.cub.2011.11.028); pmid: 22177904
- R. J. Johnston Jr., C. Desplan, Interchromosomal communication coordinates intrinsically stochastic expression between alleles. *Science* **343**, 661–665 (2014). doi: [10.1126/science.1243039](https://doi.org/10.1126/science.1243039); pmid: 24503853
- B. S. Wells, D. Pistillo, E. Barnhart, C. Desplan, Parallel Activin and BMP signaling coordinates R7/R8 photoreceptor subtype pairing in the stochastic Drosophila retina. *eLife* **6**, e25301 (2017). doi: [10.7554/eLife.25301](https://doi.org/10.7554/eLife.25301); pmid: 28853393
- K. F. Fischbach, A. P. M. Dittich, The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. *Cell Tissue Res.* **258**, 441–475 (1989). doi: [10.1007/BF00218858](https://doi.org/10.1007/BF00218858)
- N. Konstantinides et al., Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. *Cell* **174**, 622–635.e13 (2018). doi: [10.1016/j.cell.2018.05.021](https://doi.org/10.1016/j.cell.2018.05.021); pmid: 29909983
- C. Bertet, in *Decoding Neural Circuit Structure and Function: Cellular Dissection Using Genetic Model Organisms*, A. Çelik, M. F. Wernet, Eds. (Springer, 2017), pp. 419–435. doi: [10.1007/978-3-319-57363-2_17](https://doi.org/10.1007/978-3-319-57363-2_17)
- B. S. Egger, J. Q. Boone, N. R. Stevens, A. H. Brand, C. Q. Doe, Regulation of spindle orientation and neural stem cell fate in the Drosophila optic lobe. *Neural Dev.* **2**, 1 (2007). doi: [10.1186/1749-8104-2-1](https://doi.org/10.1186/1749-8104-2-1); pmid: 17207270
- K. T. Ngo, I. Andrade, V. Hartenstein, Spatio-temporal pattern of neuronal differentiation in the Drosophila visual system: A user's guide to the dynamic morphology of the developing optic lobe. *Dev. Biol.* **428**, 1–24 (2017). doi: [10.1016/j.ydbio.2017.05.008](https://doi.org/10.1016/j.ydbio.2017.05.008); pmid: 28533086
- A. Hofbauer, J. A. Campos-Ortega, Proliferation pattern and early differentiation of the optic lobes in Drosophila melanogaster. *Roux's Arch. Dev. Biol.* **198**, 264–274 (1990). doi: [10.1007/BF00377393](https://doi.org/10.1007/BF00377393); pmid: 28305665
- D. F. Ready, T. E. Hanson, S. Benzer, Development of the Drosophila retina, a neurocrystalline lattice. *Dev. Biol.* **53**, 217–240 (1976). doi: [10.1016/0012-1606\(76\)90225-6](https://doi.org/10.1016/0012-1606(76)90225-6); pmid: 825400
- X. Li et al., Temporal patterning of Drosophila medulla neuroblasts controls neural fates. *Nature* **498**, 456–462 (2013). doi: [10.1038/nature12319](https://doi.org/10.1038/nature12319); pmid: 23783517
- T. Suzuki, M. Kaido, R. Takayama, M. Sato, A temporal mechanism that produces neuronal diversity in the Drosophila visual center. *Dev. Biol.* **380**, 12–24 (2013). doi: [10.1016/j.ydbio.2013.05.002](https://doi.org/10.1016/j.ydbio.2013.05.002); pmid: 23665475
- I. Holguera, C. Desplan, Neuronal specification in space and time. *Science* **362**, 176–180 (2018). doi: [10.1126/science.aas9435](https://doi.org/10.1126/science.aas9435); pmid: 30309944
- T. Erlik et al., Integration of temporal and spatial patterning generates neural diversity. *Nature* **541**, 365–370 (2017). doi: [10.1038/nature20794](https://doi.org/10.1038/nature20794); pmid: 28077877
- M. Courgeon, C. Desplan, Coordination of neural patterning in the Drosophila visual system. *Curr. Opin. Neurobiol.* **56**, 153–159 (2019). doi: [10.1016/j.conb.2019.01.024](https://doi.org/10.1016/j.conb.2019.01.024); pmid: 30849690
- R. A. Carrillo et al., Control of Synaptic Connectivity by a Network of Drosophila IgSF Cell Surface Proteins. *Cell* **163**, 1770–1782 (2015). doi: [10.1016/j.cell.2015.11.022](https://doi.org/10.1016/j.cell.2015.11.022); pmid: 26687361
- S. Gao et al., The neural substrate of spectral preference in Drosophila. *Neuron* **60**, 328–342 (2008). doi: [10.1016/j.neuron.2008.08.010](https://doi.org/10.1016/j.neuron.2008.08.010); pmid: 18957224
- T. Karuppururai et al., A hard-wired glutamatergic circuit pools and relays UV signals to mediate spectral preference in Drosophila. *Neuron* **81**, 603–615 (2014). doi: [10.1016/j.neuron.2013.12.010](https://doi.org/10.1016/j.neuron.2013.12.010); pmid: 24507194
- L. Tan et al., Ig Superfamily Ligand and Receptor Pairs Expressed in Synaptic Partners in Drosophila. *Cell* **163**, 1756–1769 (2015). doi: [10.1016/j.cell.2015.11.021](https://doi.org/10.1016/j.cell.2015.11.021); pmid: 26687360
- K. J. Tan, et al., MiMIC: A highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nat. Methods* **8**, 737–743 (2011). doi: [10.1038/nmeth.1662](https://doi.org/10.1038/nmeth.1662); pmid: 21985007
- C. Y. Ting et al., Photoreceptor-derived activin promotes dendritic termination and restricts the receptive fields of first-order interneurons in Drosophila. *Neuron* **81**, 830–846 (2014). doi: [10.1016/j.neuron.2013.12.012](https://doi.org/10.1016/j.neuron.2013.12.012); pmid: 24462039
- S. Y. Takemura et al., A visual motion detection circuit suggested by Drosophila connectomics. *Nature* **500**, 175–181 (2013). doi: [10.1038/nature12450](https://doi.org/10.1038/nature12450); pmid: 23925240
- B. D. Pfeiffer et al., Refinement of tools for targeted gene expression in Drosophila. *Genetics* **186**, 735–755 (2010). doi: [10.1534/genetics.110.119917](https://doi.org/10.1534/genetics.110.119917); pmid: 20697123
- G. Sancer et al., Modality-Specific Circuits for Skylight Orientation in the Fly Visual System. *Curr. Biol.* **29**, 2812–2825.e4 (2019). doi: [10.1016/j.cub.2019.07.020](https://doi.org/10.1016/j.cub.2019.07.020); pmid: 31402302
- C. Bertet et al., Temporal patterning of neuroblasts controls Notch-mediated cell survival through regulation of Hid or Reaper. *Cell* **158**, 1173–1186 (2014). doi: [10.1016/j.cell.2014.07.045](https://doi.org/10.1016/j.cell.2014.07.045); pmid: 25171415
- E. Özkan et al., An extracellular interactome of immunoglobulin and LRR proteins reveals receptor-ligand networks. *Cell* **154**, 228–239 (2013). doi: [10.1016/j.cell.2013.06.006](https://doi.org/10.1016/j.cell.2013.06.006); pmid: 23827685
- S. Xu et al., Interactions between the Ig-Superfamily Proteins DIP-α and Dpr6/10 Regulate Assembly of Neural Circuits. *Neuron* **100**, 1369–1384.e6 (2018). doi: [10.1016/j.neuron.2018.11.001](https://doi.org/10.1016/j.neuron.2018.11.001); pmid: 30467079
- S. Cheng et al., Molecular basis of synaptic specificity by immunoglobulin superfamily receptors in Drosophila. *eLife* **8**, e41028 (2019). doi: [10.7554/eLife.41028](https://doi.org/10.7554/eLife.41028); pmid: 30688651
- W. B. Grueber, A. Sagasti, Self-avoidance and tiling: Mechanisms of dendrite and axon spacing. *Cold Spring Harb. Perspect. Biol.* **2**, a001750 (2010). doi: [10.1101/cshperspect.a001750](https://doi.org/10.1101/cshperspect.a001750); pmid: 20573716
- C. Anderson et al., Natural variation in stochastic photoreceptor specification and color preference in Drosophila. *eLife* **6**, e29593 (2017). doi: [10.7554/eLife.29593](https://doi.org/10.7554/eLife.29593); pmid: 29251595
- R. W. Sperry, Chemoaffinity in the Orderly Growth of Nerve Fiber Patterns and Connections. *Proc. Natl. Acad. Sci. U.S.A.* **50**, 703–710 (1963). doi: [10.1073/pnas.50.4.703](https://doi.org/10.1073/pnas.50.4.703); pmid: 14077501
- S. Barish et al., Combinations of DIPs and Dprs control organization of olfactory receptor neuron terminals in Drosophila. *PLOS Genet.* **14**, e1007560 (2018). doi: [10.1371/journal.pgen.1007560](https://doi.org/10.1371/journal.pgen.1007560); pmid: 30102700
- L. Venkatasubramanian et al., Stereotyped terminal axon branching of leg motor neurons mediated by IgSF proteins DIP-α and Dpr10. *eLife* **8**, e42692 (2019). doi: [10.7554/eLife.42692](https://doi.org/10.7554/eLife.42692); pmid: 30714901
- J. Ashley et al., Transsynaptic interactions between IgSF proteins DIP-α and Dpr10 are required for motor neuron targeting specificity. *eLife* **8**, e42690 (2019). doi: [10.7554/eLife.42690](https://doi.org/10.7554/eLife.42690); pmid: 30714906
- C. Xu et al., Control of Synaptic Specificity by Establishing a Relative Preference for Synaptic Partners. *Neuron* **103**, 865–877.e7 (2019). pmid: 31300277
- W. M. Cowan, J. W. Fawcett, D. D. O'Leary, B. B. Stanfield, Regressive events in neurogenesis. *Science* **225**, 1258–1265 (1984). doi: [10.1126/science.6474175](https://doi.org/10.1126/science.6474175); pmid: 6474175
- V. Hamburger, R. Levi-Montalcini, Proliferation, differentiation and degeneration in the spinal ganglia of the chick embryo under normal and experimental conditions. *J. Exp. Zool.* **111**, 457–501 (1949). doi: [10.1002/jez.1401110308](https://doi.org/10.1002/jez.1401110308); pmid: 18142378
- S. Cohen, Purification of a Nerve-Growth Promoting Protein From the Mouse Salivary Gland and Its Neuro-Cytotoxic Antiserum. *Proc. Natl. Acad. Sci. U.S.A.* **46**, 302–311 (1960). doi: [10.1073/pnas.46.3.302](https://doi.org/10.1073/pnas.46.3.302); pmid: 16578483

ACKNOWLEDGMENTS

We thank the fly community, the Bloomington and Kyoto Stock centers, and the DGRC for sharing flies and reagents; M. Wernet, L. Venkatasubramanian, K. Menon, and K. Zinn for sharing data prior to publication; B. Johnston and L. Zipursky for unpublished data and fly stocks; C.-H. Lee, L. Venkatasubramanian, R. Mann, and H. Bellen for fly strains; D. Godt and Y.-N. Jan for antibodies; J. Treisman, E. Mazzoni, and N. Ringstad for their input during the investigation; and all Desplan lab members for their discussion and comments on the manuscript. **Funding:** Supported by NIH grant R01 EY13010 (C.D.). **Author contributions:** M.C. and C.D., conceptualization; M.C., investigation and writing (original draft); M.C. and C.D., writing (review and editing). **Competing interests:** Authors declare no competing interests; **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/full/science.eaay6727/DC1
Figs. S1 to S6
Tables S1 and S2

View/request a protocol for this paper from Bio-protocol.

12 July 2019; accepted 18 September 2019
Published online 3 October 2019
10.1126/science.aay6727

RESEARCH ARTICLE

CELL BIOLOGY

Light-regulated collective contractility in a multicellular choanoflagellate

Thibaut Brunet^{1*}, Ben T. Larson^{1,2*}, Tess A. Linden^{1*}, Mark J. A. Vermeij³, Kent McDonald⁴, Nicole King^{1†}

Collective cell contractions that generate global tissue deformations are a signature feature of animal movement and morphogenesis. However, the origin of collective contractility in animals remains unclear. While surveying the Caribbean island of Curaçao for choanoflagellates, the closest living relatives of animals, we isolated a previously undescribed species (here named *Choanoeca flexa* sp. nov.) that forms multicellular cup-shaped colonies. The colonies rapidly invert their curvature in response to changing light levels, which they detect through a rhodopsin–cyclic guanosine monophosphate pathway. Inversion requires actomyosin-mediated apical contractility and allows alternation between feeding and swimming behavior. *C. flexa* thus rapidly converts sensory inputs directly into multicellular contractions. These findings may inform reconstructions of hypothesized animal ancestors that existed before the evolution of specialized sensory and contractile cells.

The evolution of animals from their single-celled ancestors involved several major evolutionary innovations, including multicellularity, spatial cell differentiation, and morphogenesis (1, 2). Efforts to reconstruct the origin of animal multicellularity have benefited from the study of choanoflagellates, the closest living relatives of animals (3–5). Choanoflagellates are microbial eukaryotes that feed on bacteria and live in aquatic environments around the world; many species differentiate over their life history into diverse cell types, including unicellular and multicellular forms (3, 6–8). Comparative genomics has revealed many gene families once thought to be unique to animals (such as cadherins, C-type lectins, and receptor tyrosine kinases) in choanoflagellates (4, 9–12). Moreover, the model choanoflagellate *Salpingoeca rosetta* (6) exhibits diverse responses to environmental cues, including pH-taxis (13), aerotaxis (14), and bacterial regulation of multicellular development (15) and mating (16). However, *S. rosetta* is only one of ~380 known species (17), and choanoflagellates are at least as genetically diverse as animals (9). Choanoflagellate diversity thus represents a largely untapped opportunity to investigate environmental regulation of cell behavior, the principles that broadly underpin multicellularity, and the evolution of animal cell biology. We report here on

the discovery of light-regulated collective cell contractility in a recently isolated multicellular choanoflagellate. This finding reveals that apical cell contractility evolved before the origin of animal multicellularity.

Photoc cues induce sheet inversion in a colonial choanoflagellate

During a survey of choanoflagellate diversity on the Caribbean island of Curaçao (Fig. 1, A and B), we collected large, cup-shaped colonies of protozoa (~100 μ m in diameter) from shallow splash pools above the tide line of a rocky coastal area (Fig. 1B). Each colony was composed of a monolayer (“sheet”) of up to hundreds of flagellated cells (Fig. 1C and movie S1). The cells bore the characteristic collar complex of choanoflagellates (1, 3), in which a “collar” of microvilli surrounds a single apical flagellum (Fig. 1D). However, unlike in most choanoflagellate colonies (3), the apical flagella pointed into the interior of the colony (Fig. 1E), resembling the orientation of collar cells in the choanocyte chambers of sponges (18). The colonies inverted rapidly (within ~30 s from initiation to completion) while maintaining their cell topology, such that the flagella pointed outward along the radius of curvature of the colony (Fig. 1, F and G, and movie S2). The colonies tended to remain in the inverted form (“flagella-out”) for several minutes before reverting to their initial, relaxed conformation (“flagella-in”) in a similarly rapid process (Fig. 1H and movie S3).

To start laboratory cultures of this choanoflagellate, we manually isolated representative colonies away from the other microbial eukaryotes in the splash pool sample (for example, presumptive *Oxyrrhis* sp. dinoflagellates) (Fig. 1C and movie S1) and transferred them into

nutrient-supplemented artificial seawater along with co-isolated environmental bacteria (which choanoflagellates need as a food source). Analysis of the 18S ribosomal DNA (rDNA) sequence indicated that the choanoflagellate is the sister species of the previously described *Choanoeca perplexa* (also known as *Proterospongia choanojuncta*) (Fig. 1J and fig. S1), whose life history includes both single cells and colonies (7, 19). Inspired by the sheet-bending behavior of the Curaçao choanoflagellate, we named it *Choanoeca flexa*. [A similar behavior was mentioned in a 1983 study of *C. perplexa* (7), but the culture subsequently stopped forming colonies, preventing further study (3).] *C. flexa* sheet inversion, representing a global change in multicellular form, is reminiscent of concerted movement and morphogenesis in animals (such as muscle contraction or gastrulation). Because of the potential evolutionary implications of rapid shape change in *C. flexa*, we investigated (i) how colony inversion is regulated, (ii) the mechanisms underlying colony inversion, and (iii) the ecological consequences of colony inversion.

Several lines of evidence indicated that light levels regulate *C. flexa* colony inversion. While imaging *C. flexa* sheets for >1 hour under constant illumination, colony inversions became less frequent. By contrast, after the microscope illumination was turned off, the colonies inverted almost immediately (movie S4). To quantify light-to-dark-induced *C. flexa* inversion, we established an assay based on the observation that the projected area of a *C. flexa* sheet decreases by as much as 50% during inversion (Fig. 2, A to C, and movie S5). Using this assay, we confirmed that a rapid decrease in illumination reliably induces inversion of *C. flexa* colonies within 30 s (Fig. 2D and movie S6). Choanoflagellates had previously been thought to be insensitive to light (20).

A rhodopsin-cGMP pathway regulates colony inversion in response to light-to-dark transitions

We next investigated how *C. flexa* colonies sense and respond to changing light levels. Although choanoflagellates are unpigmented and transparent, at least four species encode a choanoflagellate-specific rhodopsin-phosphodiesterase fusion protein (fig. S2) (9), RhoPDE, that has been investigated as a potential optogenetic tool (21–24). RhoPDEs consist of an N-terminal type I rhodopsin [a photosensitive transmembrane protein broadly involved in light detection (25)] fused to a C-terminal phosphodiesterase (PDE) that catalyzes cyclic nucleotide hydrolysis (Fig. 3A). In *in vitro* studies (21–24), *S. rosetta* RhoPDE appears capable of converting a photic stimulus into a biochemical signal within seconds, similar to the time scale of the *C. flexa* response to light-to-dark transitions.

¹Howard Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. ²Biophysics Graduate Group, University of California, Berkeley, CA, USA. ³Department of Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, CARMABI, Piscaderabaai z/n Willemstad, Curaçao. ⁴Electron Microscopy Laboratory, University of California, Berkeley, CA, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: nking@berkeley.edu

To search for RhoPDE or other candidate photosensitive proteins in *C. flexa*, we sequenced and assembled the *C. flexa* transcriptome (26), which we found to encode four RhoPDE homologs (fig. S2) (GenBank accession numbers MN013138, MN013139, MN013140, and MN013141). No other rhodopsins were detected in the *C. flexa* transcriptome. The only other candidate photoreceptor

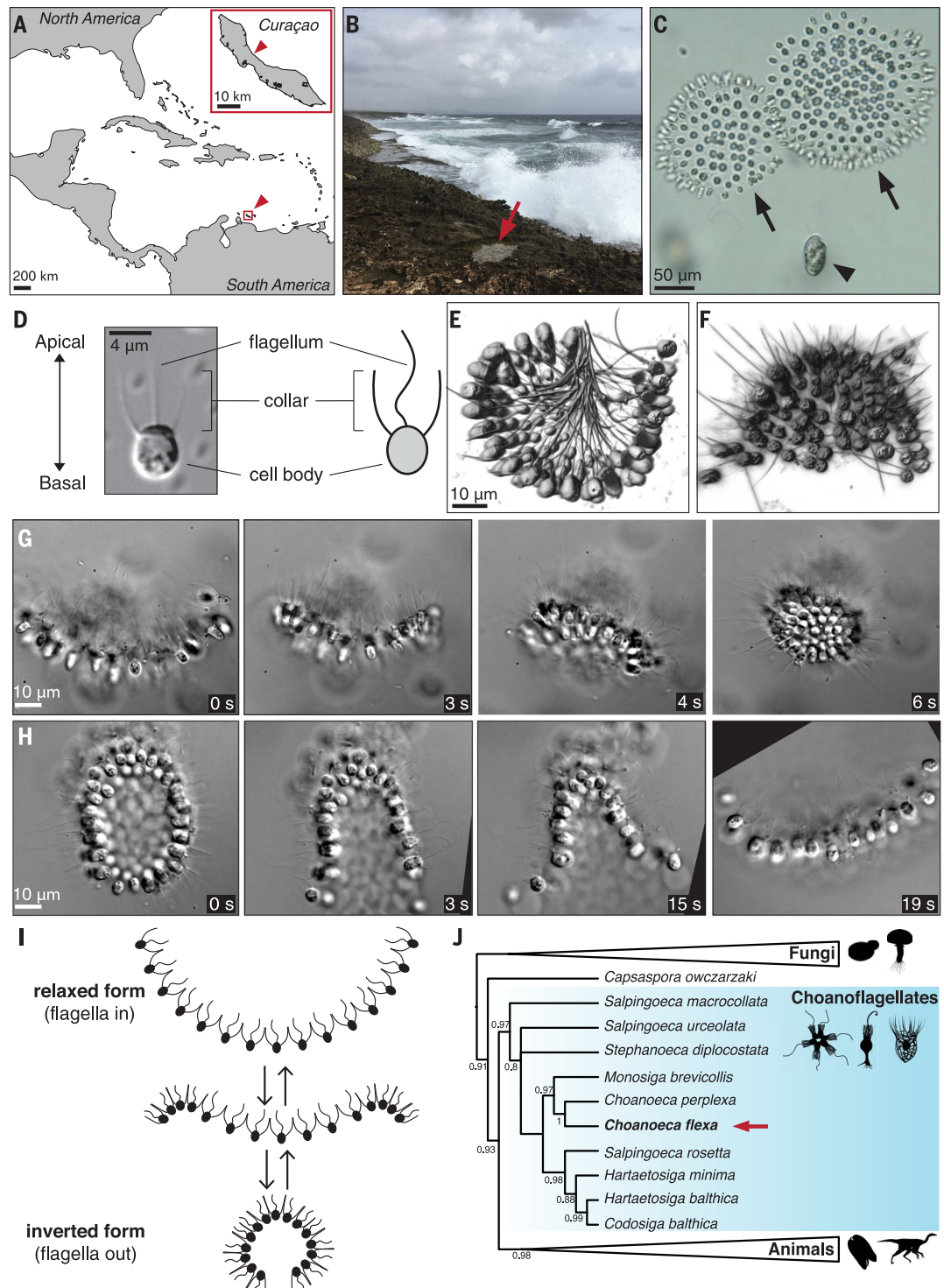
protein domain found was a cryptochrome transcription factor (27), which is predicted to act on the time scale of transcriptional regulation [at least several minutes (28, 29)] and, therefore, is unlikely to mediate the light-to-dark transition response. Thus, we focused our attention on the RhoPDEs.

If RhoPDE regulates the light-to-dark transition response, depletion of the rhodopsin

chromophore, retinal (30, 31), should abolish rhodopsin activity and thereby prevent the response. Moreover, artificially increasing the cellular concentration of cyclic guanosine monophosphate (cGMP) or cyclic adenosine monophosphate (cAMP) (which are degraded by the enzymatic activity of PDEs) should mimic the effect of darkness and be sufficient to trigger sheet inversion.

Fig. 1. Multicellular sheets of a colonial choanoflagellate from Curaçao rapidly and reversibly invert in curvature.

(A to C) *C. flexa* was isolated from splash pools on the northern shore of Curaçao. (A) Map of the Caribbean Sea. Arrowhead: Curaçao. (Inset) Map of Curaçao. Inset arrowhead: sampling site (12°13'38.9" N 69°00'47.0" W). (B) Photograph of the sampling site. Arrow: splash pool. (C) Microscopy of freshly collected splash pool sample revealed a diverse microeukaryotic community, including dinoflagellates (*Oxyrrhis* sp.; arrowhead) and cup-shaped choanoflagellate colonies (*C. flexa*; arrows) that rapidly inverted their curvature. Still frame from movie S1. (D) Diagnostic features of choanoflagellate cell morphology shown by differential interference contrast (DIC) micrograph and sketch of a *C. flexa* cell. (E and F) *C. flexa* colonies alternate between two conformations, flagella-in (E) and flagella-out (F). (G and H) *C. flexa* colonies rapidly and reversibly invert their curvature while maintaining contacts among neighboring cells. (G) Flagella-in colony inverts to the flagella-out orientation (movie S2). (H) Flagella-out colony reverts to the flagella-in orientation (movie S3). Movie frames were rotated to facilitate tracking individual cells between images. (I) Summary of the inversion and relaxation processes. (J) Phylogenetic analysis of 18S rDNA sequences revealed that *C. flexa* (bold) is a sister to the choanoflagellate *C. perplexa* (19). (Scaled branch lengths are in Fig. S1.)



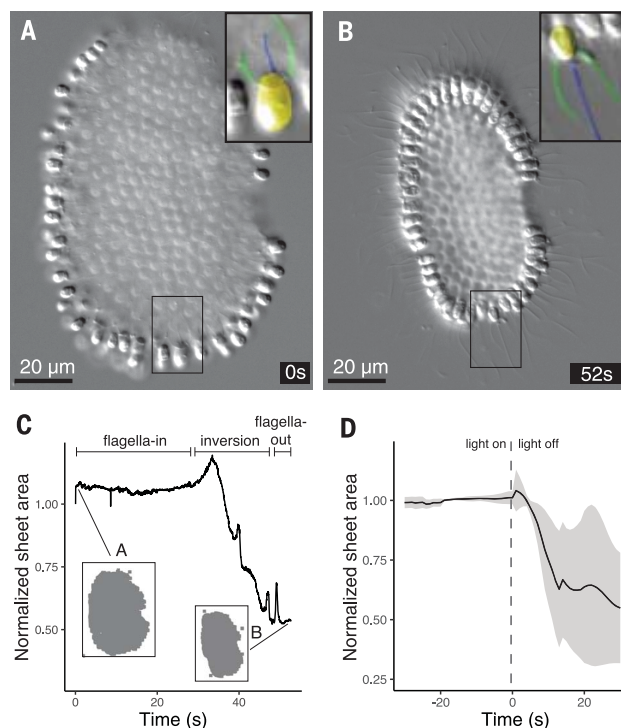


Fig. 2. Light-to-dark transitions induce *C. flexa* colony inversion. (A to C) Colony inversion correlates with a decrease in the projected area of the colony. (A and B) *C. flexa* colony spontaneously inverts from the flagella-in (A) to the flagella-out (B) orientation (movie S5). (Insets) Pseudocoloring highlights the orientations of the cells in the boxed regions. Cell orientation relative to colony curvature inverts without breaking contacts with neighboring cells. (C) The projected area of the colony from (A) and (B) decreased by ~50% during inversion. Sheet area was normalized to its preinversion projected area. (D) Colonies reliably undergo inversion in response to light-to-dark transitions. Normalized projected area of $n = 5$ colonies over time before and after light reduction (vertical dashed line). Line: mean projected area (rolling average over 5-s windows); ribbon: standard deviation.

Plants and some bacteria synthesize retinal and other carotenoids, but the *C. flexa* transcriptome lacks a key enzyme in the retinal biosynthesis pathway (fig. S3). Therefore, like animals, *C. flexa* must receive retinal or its precursor, beta-carotene, from its food. It is possible that *C. flexa* acquires retinal or beta-carotene from its bacterial prey, with which it is cultured. To test whether bacterially produced carotenoids are required for light-regulated colony inversion, we established a culture containing only *C. flexa* and a co-isolated bacterium, *Pseudomonas oceani*, that lacks genes in the retinal biosynthesis pathway (fig. S3) (32). This culture, named “ChoPs” (for *Choanoeca* + *Pseudomonas*), was expected to be devoid of carotenoids, thereby abolishing rhodopsin activity. As predicted, *C. flexa* sheets in ChoPs cultures did not invert in response to darkness (Fig. 3B). Inoculating ChoPs cultures with a mixture of co-isolated environmental bacteria restored the light-to-dark response, demonstrating that a bacterial factor is necessary for the response. Addition of exogenous retinal to ChoPs cultures restored the wild-type light-to-dark response (Fig. 3C), indicat-

ing that the absence of retinal explains the absence of colony inversion. The requirement of retinal for the inversion response and the fact that all rhodopsin-encoding genes in the *C. flexa* are RhoPDEs suggest that one (or more) RhoPDE mediates the light-to-dark-induced inversion.

We next investigated whether cyclic nucleotide signaling influences *C. flexa* inversion. Treatment of *C. flexa* sheets with two inhibitors of PDE activity, caffeine (33) and 3-isobutyl-1-methylxanthine (IBMX) (34), induced colony inversion in the absence of a photic stimulus (Fig. 3D). Moreover, incubating *C. flexa* sheets with a cell-permeant analog of cGMP induced colony inversion in a dose-responsive manner, whereas a cell-permeant analog of cAMP had no effect (Fig. 3E), suggesting that cGMP acts as a second messenger in phototransduction and thereby triggers colony inversion.

Together, these results indicate that the *C. flexa* response to light-to-dark transitions relies on a rhodopsin as a photoreceptor and cGMP as a second messenger. The simplest interpretation of these findings is that a RhoPDE controls *C. flexa* phototransduction.

However, direct validation will require targeted disruption of all four *C. flexa* RhoPDE homologs (fig. S2).

Sheet inversion mediates a trade-off between feeding and swimming

What are the functional and ecological roles of sheet inversion in *C. flexa*? Flagella-in sheets showed little to no motility, either slowly drifting or settling to the bottom of flasks (movie S7), whereas inverted (flagella-out) sheets swam rapidly (Fig. 4, A and B; fig. S5; and movie S4). By contrast, cells in flagella-in sheets fed efficiently (>75% cells per sheet internalizing beads), whereas cells in flagella-out sheets did not (~10% cells per sheet internalizing beads on average) (Fig. 4, C to G). Moreover, in relaxed sheets, fluid flow converged toward the center of the colony (carrying bacterial prey toward the cells), whereas in inverted sheets, the flow was directed away from the colony, allowing swimming but not feeding (fig. S6). Thus, sheet inversion increases motility, which may allow escape from environmental hazards (including predators), whereas sheet relaxation allows enhanced feeding efficiency.

Because sheets swim slowly when relaxed and rapidly when inverted, we suspected that darkness-induced inversion might allow sheets to accumulate in bright areas, effectively undergoing phototaxis. Using chambers illuminated with directional light, we found that *C. flexa* sheets tended to accumulate near the illumination source compared with a control in which no illumination was provided (Fig. 4H and fig. S7). Further, neither sheets from ChoPs cultures nor single cells from dissociated sheets were capable of phototaxis, suggesting that rhodopsin activity, multicellularity, and sheet inversion are all required for phototaxis (Fig. 4H). Thus, sheet inversion mediates an ecologically relevant trade-off between feeding and swimming (Fig. 4I).

Sheet inversion requires apical actomyosin contractility

How do cells in sheets interact, and what mechanisms allow sheet inversion? We found that cells in *C. flexa* sheets form direct contacts between their collar microvilli (Fig. 5, A to C, and fig. S8) but not through the intercellular bridges, shared extracellular matrix, or filopodial contacts that mediate multicellularity in other choanoflagellate species (6, 7, 10, 35). Collar morphology differs between relaxed and inverted sheets. In relaxed sheets (flagella-in; Fig. 5C), the microvilli on each cell assemble into a barrel-shaped collar whose diameter varied little from base to tip. In inverted sheets (flagella-out; Fig. 5C), the microvilli form a flared, cone-shaped collar whose diameter increases from base to tip. Active “opening out” of the collar, by increasing

the surface area of the apical side of the sheets relative to their basal side, might force a change in sheet curvature. Consistent with this, the collar of dissociated *C. flexa* cells treated with caffeine opened into a conical shape, straightened, and slid down toward the base of the cell, whereas untreated controls maintained a barrel-shaped collar (Fig. 5, D to G, and fig. S9). These data suggest that the changes in collar geometry during inversion are actively generated by individual cells and do not require interactions among neighboring cells.

Sheet bending in animal epithelia is frequently mediated by apical constriction, in which contraction of an apical actomyosin network reduces the surface area of the apical side of the cell (36, 37). Apical constriction is

mediated by molecular motors belonging to the myosin II family, which predates the diversification of modern eukaryotes (38) and is found in all choanoflagellate genomes (39) and transcriptomes (9) published to date. *C. flexa* encodes homologs of the myosin II regulatory light chain (GenBank accession MK787241) and heavy chain (GenBank accession MK787240) (fig. S4), whose protein sequences are 78 and 63% similar, respectively, to their human counterparts.

In choanoflagellates and animal epithelial cells, the apical pole is defined by the presence of a flagellum or cilium, respectively, and/or microvilli, and the apicobasal axis of both types of cells is broadly considered to be homologous (40). *C. flexa* sheets contain a pronounced F-actin

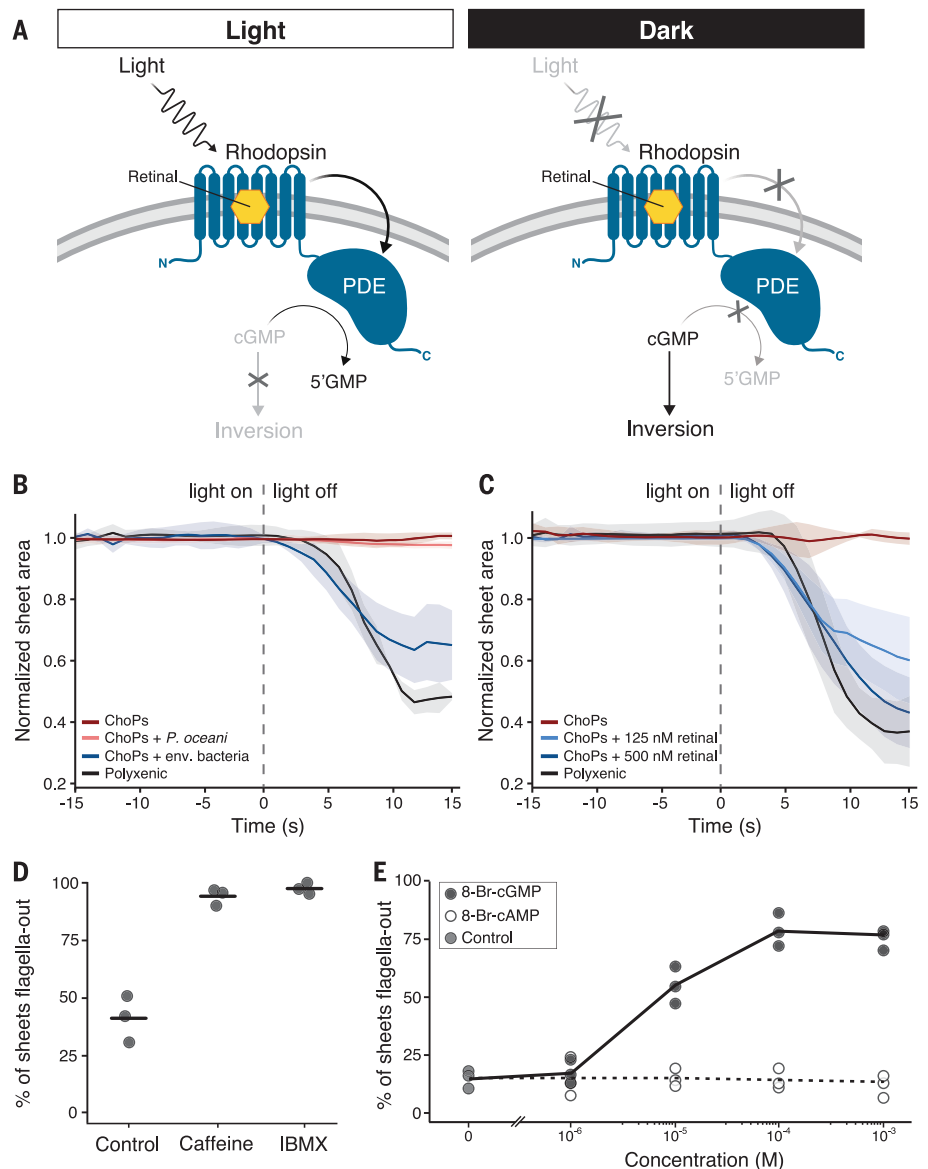
ring at the apical pole of each cell, from which the microvillar collar extends (Fig. 5H). Extending from this ring, we detected a small number of longitudinal actin fibers (usually two or three) pointing toward the basal pole. Diameter measurements showed that the F-actin ring was consistently smaller in inverted sheets compared with relaxed sheets (Fig. 5, I to L). The same was true for dissociated, caffeine-treated cells compared with the corresponding negative controls (Fig. 5, I to J and L), consistent with the ring actively constricting during sheet inversion. During inversion and in response to caffeine, some (but not all) cells transiently acquired a “bottle cell” morphology with a narrow apex and a bulbous base (fig. S9), reminiscent of animal

Fig. 3. *C. flexa* cells transduce light stimuli through a rhodopsin-cGMP pathway using bacterial carotenoids. (A) RhoPDE (blue), a choanoflagellate-specific enzyme rhodopsin.

C. flexa encodes four copies of RhoPDE (fig. S2), each comprising an 8-transmembrane-pass type I (bacterial) rhodopsin fused to a cyclic nucleotide PDE. Photodetection by rhodopsin requires binding of the chromophore retinal. Under illumination (left panel), the rhodopsin domain activates the PDE domain, which hydrolyzes cGMP to 5'GMP, thereby reducing cellular cGMP levels. Upon light reduction (right panel), the PDE domain is inactivated, allowing cellular cGMP levels to rise. **(B)** A bacterial factor is required for light-regulated sheet inversion.

In the presence of diverse “polyxenic” bacteria, *C. flexa* sheets inverted to flagella-out in response to decreased illumination ($n = 3$ colonies). By contrast, *C. flexa* sheets grown only with the bacterium *P. oceanii* (ChoPs culture, $n = 4$ colonies) did not respond to changes in illumination. The photic response of the ChoPs culture was restored by inoculating with environmental bacteria from the polyxenic culture (ChoPs + env. bacteria, $n = 3$ colonies) but not by inoculating with *P. oceanii* bacteria (ChoPs + *P. oceanii*, $n = 3$ colonies). **(C)** Retinal (or one of its carotenoid precursors) is the bacterial molecule required for the photic response. The photic response of the ChoPs culture, which normally does not respond to light stimuli ($n = 4$ colonies), was restored by treatment with 125 nM or 500 nM retinal ($n = 5$ colonies each). Thus, retinal is required for inversion in response to reduced illumination. For (B) and (C), photic response was quantified as for Fig. 2.

(D) PDE activity suppresses sheet inversion in *C. flexa*. Treatment with the PDE inhibitors caffeine (10 mM) or IBMX (1 mM) caused *C. flexa* colonies to invert to the flagella-out orientation in the absence of a photic stimulus ($n = 3$ independent trials; $N = 52, 55$, and 38 sheets for controls; $N = 23, 31$, and 40 sheets for caffeine; $N = 42, 37$, and 27 sheets for IBMX). **(E)** cGMP induces sheet inversion. Treating the ChoPs culture with a cell-permeant cGMP analog (8-Br-cGMP) caused sheets to invert into the flagella-out orientation in a dose-dependent manner in the absence of a photic stimulus. Treating with 8-Br-cAMP had no effect.



cells undergoing pronounced apical constriction (41, 42). Caffeine treatment also induced shortening of the longitudinal actin fibers (fig. S9, G to H), suggesting that fiber contraction pulls the collar toward the basal pole.

Using an antibody raised against *C. flexa* myosin II as well as five different commercial myosin II antibodies (Fig. 5, M to S, and fig. S10), we found that *C. flexa* cells contain myosin that overlaps in regions with the apical actin ring (Fig. 5, M to P, and fig. S10) and longitudinal fibers (Fig. 5, Q to S), consistent with the idea that the apical actin network is contractile. Inhibition of myosin II activity with blebbistatin (43) abolished ring constriction in caffeine-treated dissociated cells (fig. S11A) and prevented sheet inversion (Fig. 5T), as did inhibition of dynamic actin

polymerization with latrunculin B (44) and inhibition of phosphorylation of the myosin regulatory light chain with ML-7 (45) (Fig. 5T). None of these drugs affected flagellar beating, consistent with them specifically targeting actomyosin. Although compaction is normally associated with inversion, drug treatment resulted in sheet compaction in the absence of inversion (fig. S11B), suggesting that a baseline level of tension is needed to maintain spacing between cells. Together, these results suggest that sheet inversion requires apical constriction of an actomyosin network at the base of the collar (Fig. 5U).

The ancestry of apical constriction

The discovery of sheet bending driven by apical constriction in a multicellular choanoflagel-

late has several potentially important evolutionary implications. Epithelial sheet bending is a fundamental mechanism underlying animal embryonic development (36, 46, 47), and multicellular contractility also plays a fundamental role in the behavior of adult animals by allowing fine-tuned body deformations (48). As both embryonic and adult tissue contractility are found in nearly all animal lineages, including sponges (49, 50), ctenophores (51, 52), placozoans (53), cnidarians (41, 54), and bilaterians (36, 37, 55), both were likely present in the last common animal ancestor. By contrast, collective contractility and apical constriction were hitherto unknown in close relatives of animals, making their origin mysterious.

The existence of actomyosin-mediated apical constriction in *C. flexa* raises the possibility

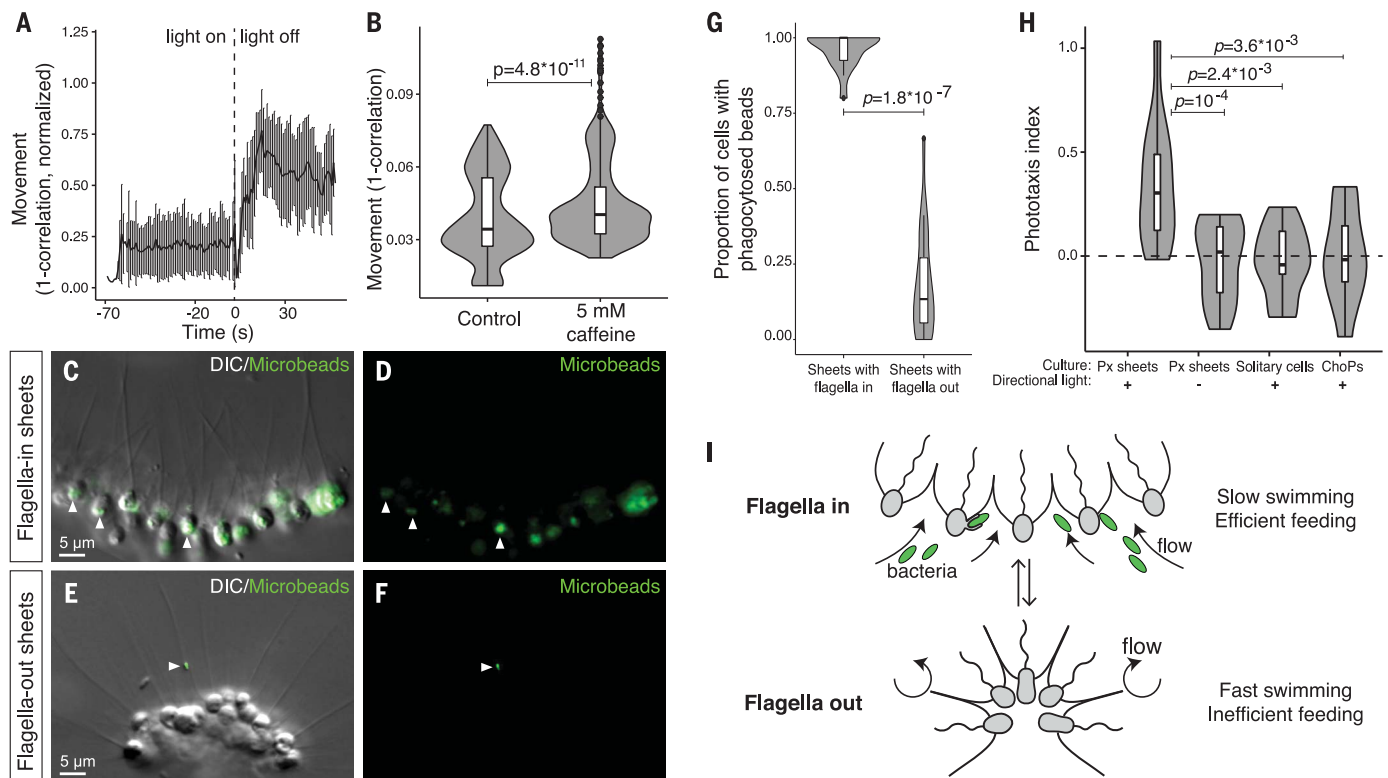


Fig. 4. Sheet inversion mediates a trade-off between swimming and feeding.

(A and B) Flagella-out sheets swim faster than flagella-in sheets. (A) After light-to-dark-induced inversion, swimming speed increased quickly (movie S4), as quantified by an increase in the amount of movement (defined as 1-correlation, where “correlation” is the Pearson correlation between two consecutive frames). Movement was normalized between 0 and 1 for each of $n = 9$ movies. Gray shading: error bars representing standard deviation. (B) Sheets swim faster after caffeine-induced inversion under constant light. Movement quantified as in (A). $n = 9$ time-lapse movies for controls (relaxed sheet populations under constant light) and $n = 10$ movies for caffeine-treated samples. (C to G) Flagella-in sheets feed more efficiently than flagella-out sheets. Internalization of 0.2- μ m fluorescent beads was used to quantify phagocytic activity. (C to F) Detection of fluorescent microbeads phagocytosed by flagella-in sheets (untreated) and flagella-out sheets (caffeine-treated), fixed after 1 hour. *C. flexa* cells were visualized by DIC (C and E) and beads by green fluorescence (C to F). Arrowheads: fluorescent beads [inside the

cells in (C); stuck to a flagellum in (E)]. (G) Proportion of cells having phagocytosed beads in $n = 17$ flagella-in sheets compared with $n = 21$ flagella-out sheets. P value: χ^2 test. (H) *C. flexa* phototaxes in a retinal- and multicellularity-dependent manner, as measured by the phototaxis index, which quantifies directional accumulation of the sheets toward a localized light source (materials and methods). Light-responsive polyxenic (Px) sheets migrated toward a lateral light source over 1 hour ($n = 12$ experiments). By contrast, no directional accumulation was observed in Px sheets without directional light ($n = 12$ experiments), in dissociated single cells ($n = 9$ experiments), or in retinal-deprived light-insensitive monoxenic cultures (ChoPs; $n = 10$ experiments) that do not undergo inversion. P values: analysis of variance with Dunnett’s correction. (I) In flagella-in sheets, flagellar beating generates a feeding flow that carries bacteria toward the basal side of the cells (fig. S7), allowing phagocytosis. In flagella-out sheets, flagellar beating causes rapid swimming, whereas the basal side of the cells faces inward, preventing prey capture.

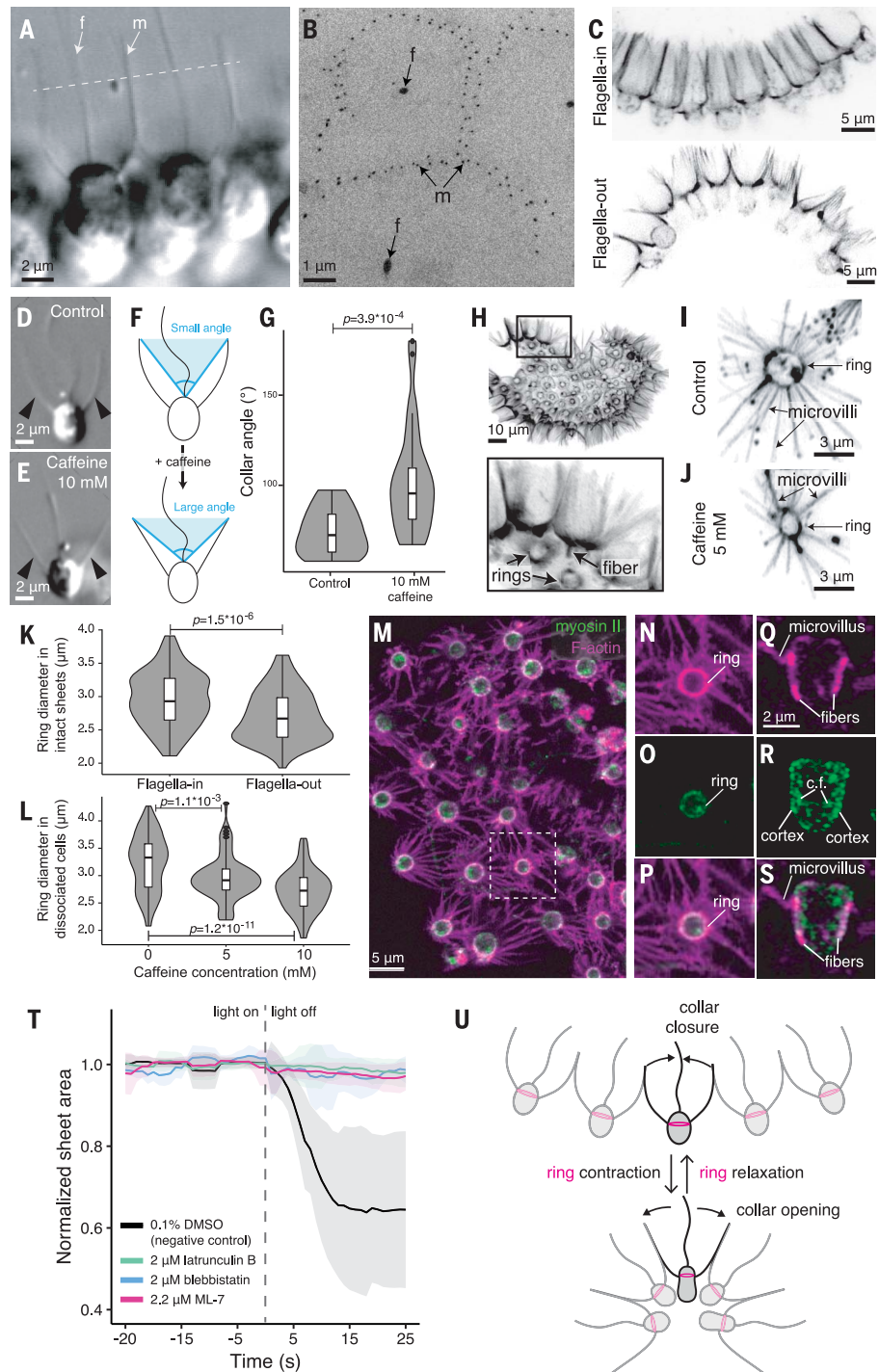
that this cellular module might have been present in the last common ancestor of choanoflagellates and animals [which together compose the choanozoans (1)]. In addition to *C. flexa*, in which collar contractions occur in sheets (Fig. 5G), dissociated cells from sheets (Fig. 5, D and E), and naturally solitary “thecate”

cells (movie S8), collar contractions have also been reported in unicellular stages from three other choanoflagellates: *Codosiga pulcherrima* (56), *Monosiga gracilis* (57), and *C. perplexa* (7). We found that four other choanoflagellate species, *Monosiga brevicollis*, *S. rosetta*, *Salpingoeca urceolata*, and *Diaphanoeca grandis*, which

together cover both main branches of the choanoflagellate phylogenetic tree (17), display spontaneous changes in collar geometry occurring at the scale of a few seconds. *S. urceolata* (movie S9) and *M. brevicollis* (movie S10) showed spontaneous and reversible opening and closing of the collar (similar to *C. flexa*),

Fig. 5. Sheet inversion requires apical actomyosin cell contraction.

(A and B) Cells in a sheet are linked by their collars. (A) DIC micrograph showing direct collar contacts between neighboring cells. f, flagellum; m, microvilli. Dotted line: approximate plane of section in (B). (B) Transmission electron micrograph of a transverse section through the collars of neighboring cells. (C) Collar morphology differs between flagella-in sheets (top) and flagella-out sheets (bottom). Cells in flagella-in sheets have barrel-shaped collars, whereas cells in flagella-out sheets have flared, conical collars. Staining: phalloidin. (D to G) Caffeine treatment of dissociated cells causes the collar to flare. (D) In the absence of caffeine, the collar of a cell from a dissociated colony curves upward. (E) In the presence of 10 mM caffeine, the collar of a cell from a dissociated colony straightened and flared open. Arrowheads: microvilli. (F) Differences in collar morphology where quantified in control and caffeine-treated single cells by measuring the collar angle (defined by the tip of two opposite microvilli and the base of the flagellum) (fig. S9). (G) Collar angles in caffeine-treated cells ($n = 28$) are wider than in control cells ($n = 16$). P value from Mann-Whitney U test. (H) Actin ring connected to longitudinal fibers is present at the base of each collar. (Inset) Higher magnification showing actin rings and a longitudinal fiber. Staining: phalloidin. (I to L) Actin ring constricts during inversion in intact sheets and in response to caffeine in isolated cells. (I and J) Actin ring observed with phalloidin in an untreated cell (I) and in a cell treated with 5 mM caffeine (J). (K) Ring diameter is larger in flagella-in sheets ($n = 8$ sheets, $N = 124$ cells) than in flagella-out sheets ($n = 7$ sheets, $N = 110$ cells). P value from Mann-Whitney U test. (L) Ring diameter is smaller in 10 mM caffeine-treated dissociated cells ($n = 74$) and 5 mM caffeine-treated dissociated cells ($n = 82$) than in untreated dissociated cells ($n = 89$). P values are from Dunnett's test for comparing several treatments with a control. (M to S) *C. flexa* myosin II localizes to the actin ring and longitudinal fibers. (M) Immunostained sheet showing apical rings of actin and myosin II at the apical poles of all cells. (N to P) Closer views showing the apical ring. (Q to S) Side view of a stained *C. flexa* cell (apical side up) showing cytoplasmic foci of myosin II (c.f.), as well as cortical staining (cortex) that overlaps with the longitudinal fibers. Green: anti-*C. flexa* myosin II antibody, magenta: rhodamine-phalloidin (fig. S10). (T) Treatment with inhibitors of actin polymerization (latrunculin B, $n = 6$), myosin contractility (blebbistatin, $n = 6$), or myosin activation by phosphorylation (ML-7, $n = 9$) prevented sheet inversion in response to light-to-dark transitions. $n = 13$ DMSO-treated controls. Photoc response is quantified as in Fig. 2. DMSO, dimethyl sulfoxide. (U) Proposed role of actomyosin contractility in inversion.



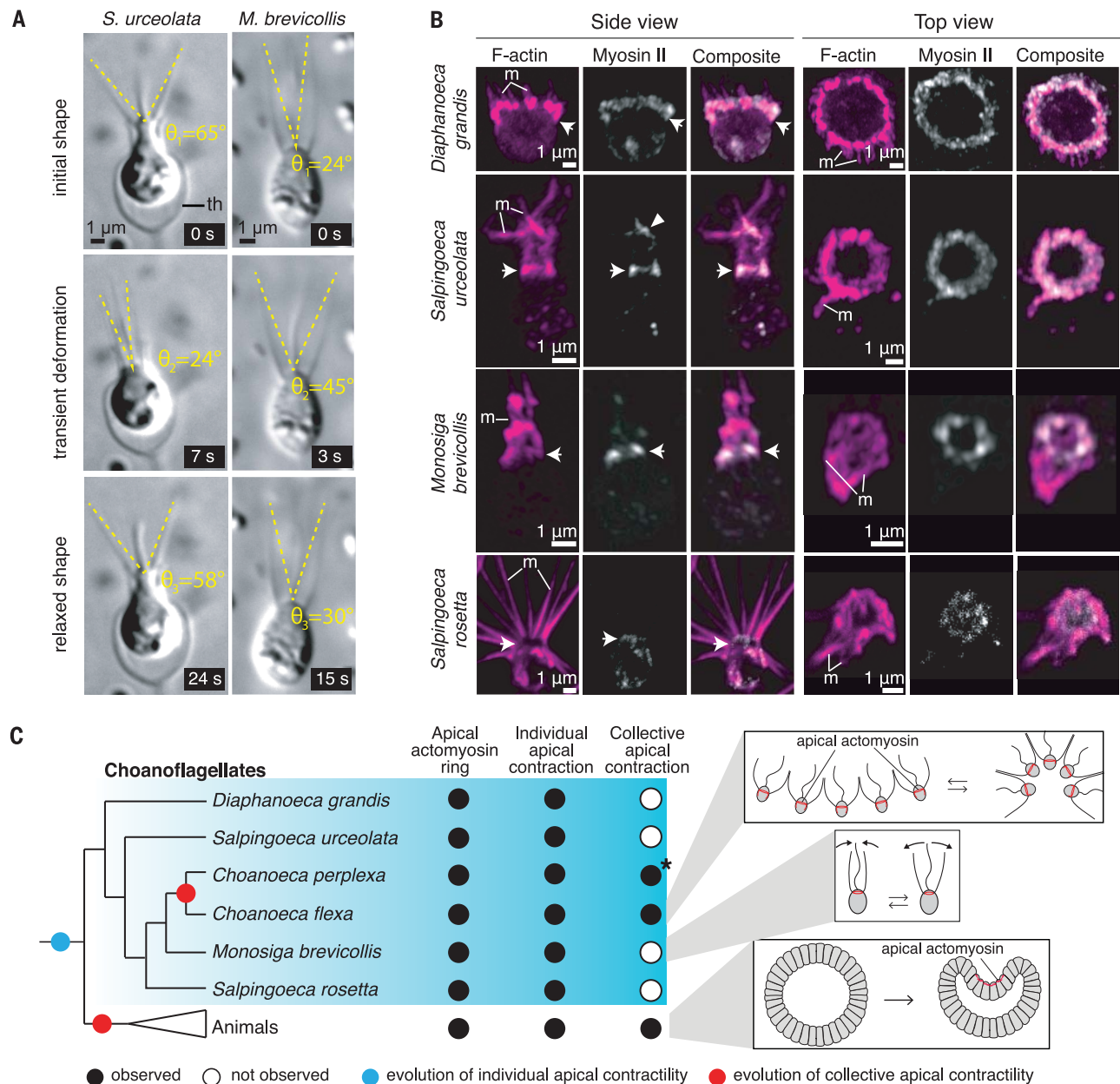


Fig. 6. Apical constriction is conserved in choanoflagellates. (A) Spontaneous collar contractions observed in *S. urceolata* and *M. brevicollis* (movies S8 and S9). Collar angles measured as in Fig. 5F. th, theca. (B) Apical actomyosin ring (arrows) is detected at the base of the collar in four representative choanoflagellate species. Myosin II was also detected in microvilli (m) in *S. urceolata* (arrowhead).

(C) Apical constriction of individual cells was present in the last common ancestor of choanozoans and independently gave rise to multicellular apical constriction in *C. flexa* and in animals. **C. perplexa* was previously reported (7) to undergo transient inversions of colony curvature. On the basis of our study of *C. flexa*, we hypothesize that these inversions reflect collective apical constriction.

whereas *S. rosetta* (movie S11) and *D. grandis* (movie S12) displayed subtler shape changes (reorientation of individual microvilli and modulation of collar curvature, respectively) (Fig. 6A and fig. S12). Like animal epithelial cells and *C. flexa*, all four species have an apical actomyosin ring at the base of the collar (Fig. 6B). This suggests that the apical actomyosin ring is a conserved feature of choanoflagellate biology (Fig. 6C) and that unicellular apical constriction was present in the last common ancestor of choanoflagellates and animals.

What is the function of apical constriction in single cells? In some sessile choanoflagellates, including in the thecate form of *C. perplexa*, collar contraction in response to mechanical stimulation allows retraction of the cell inside an extracellular structure called a theca (19), suggesting it represents a defensive withdrawal reflex from predators or other threats. In free-swimming cells, collar contraction might fine-tune the hydrodynamics of swimming or feeding. For example, a closed collar might reduce drag and facilitate locomotion,

whereas a flared collar could slow down swimming and increase collar area, thereby facilitating prey capture. Validation of these functional hypotheses will require direct testing. In contrast to single-cell apical constriction, the multicellular sheet bending observed in *C. flexa* and *C. perplexa* (7) has not been reported in other choanoflagellates. This suggests that apical constriction was present in solitary cells in the last choanozoan common ancestor and was independently converted into multicellular sheet bending through the

evolution of intercellular junctions in animals (58) and the evolution of microvillar adhesions in *C. flexa*. Multicellular inversion has been proposed to have been part of the developmental repertoire of ancient animals (1, 59), on the basis of the existence of whole-embryo inversion (from flagella-in to flagella-out) during calcareous sponge development (60). A similar inversion (but much slower, about an hour long, and irreversible) takes place during the development of the alga *Volvox* (61). Given the large evolutionary distance between choanoflagellates and volvoclean green algae, along with the absence of inversion in intervening branches, inversion likely evolved independently in both groups (1).

These observations suggest that apical actomyosin-mediated cell constriction evolved on the choanozoan stem lineage (Fig. 6B). Could polarized actomyosin contractility be even more ancient? Polarized actomyosin contractions have been implicated in multicellular morphogenesis in the fruiting body of the slime mold *Dictyostelium* (62) and may be homologous to those observed in choanoflagellates and animals. However, the absence of comparable processes in the intermediate branches between *Dictyostelium* and choanozoans raises the possibility that polarized cell contractions in *Dictyostelium* and apical constriction in choanozoans evolved independently (63). The ichthyosporean *Sphaeroforma arctica*, a close relative of choanozoans, forms large multinucleated spores that partition into distinct cells in an actomyosin-dependent process (64), providing an independent example of actomyosin-dependent multicellular development.

In animals, the control of multicellular contractions invariably relies either on the co-operation of multiple cell types [as in adult organisms (54, 55, 65)] or on complex programmed signaling cascades [as in embryos (36, 46, 47, 66, 67)]. By contrast, *C. flexa* directly converts sensory stimuli into collective contractions, without observable spatial cell differentiation, and evokes some hypotheses of early animal evolution that envisioned the first contractile tissues as homogeneous myo-epithelia of multifunctional sensory-contractile cells (68).

The fact that contractility in *C. flexa* can be controlled by light represents another intriguing parallel to animal biology. Indeed, rhodopsin-cGMP pathways similar to that in *C. flexa* also underlie phototransduction in some animal cells [for example, bilaterian ciliary photoreceptors (31, 69) and cnidarian photoreceptors (70, 71)], as well as in fungal zoospores (72). In contrast with choanoflagellates, however, phototransduction in animal photoreceptors relies on a type II (eukaryotic) rhodopsin that activates a separate PDE through a

G-protein intermediary (31, 69) (fig. S13). Meanwhile, fungal zoospores use a distinct rhodopsin fusion protein (a type I rhodopsin fused to a guanylyl cyclase) to increase cellular cGMP in response to light (72) (fig. S13). If a RhoPDE fusion protein controls *C. flexa* phototransduction, this would represent a third independent solution to the problem of transducing information from a change in illumination into a change in cyclic nucleotide signaling.

Much remains to be discovered concerning the ecological function, mechanical underpinnings, and molecular mechanisms of phototransduction and apical constriction in *C. flexa*. A deeper understanding will require the development of molecular genetic tools, which have only recently been established in *S. rosetta* (35, 73) and *D. grandis* (74). Nonetheless, *C. flexa* demonstrates how the exploration of choanoflagellate diversity can reveal biological phenomena and provides an experimentally tractable model for studying multicellular sensory-contractile coupling.

REFERENCES AND NOTES

1. T. Brunet, N. King, *Dev. Cell* **43**, 124–140 (2017).
2. A. Seb  -Pedr  s, B. M. Degnan, I. Ruiz-Trillo, *Nat. Rev. Genet.* **18**, 498–512 (2017).
3. B. S. C. Leadbeater, *The Choanoflagellates* (Cambridge University Press, 2015).
4. N. King et al., *Nature* **451**, 783–788 (2008).
5. I. Ruiz-Trillo, A. J. Roger, G. Burger, M. W. Gray, B. F. Lang, *Mol. Biol. Evol.* **25**, 664–672 (2008).
6. M. J. Dayel et al., *Dev. Biol.* **357**, 73–82 (2011).
7. B. S. C. Leadbeater, *J. Mar. Biol. Assoc. U. K.* **63**, 135–160 (1983).
8. D. Laundon, B. T. Larson, K. McDonald, N. King, P. Burkhardt, *PLOS Biol.* **17**, e3000226 (2019).
9. D. J. Richter, P. Fozouni, M. B. Eisen, N. King, *eLife* **7**, e34226 (2018).
10. T. C. Levin, A. J. Greaney, L. Wetzel, N. King, *eLife* **3**, e04070 (2014).
11. N. King, C. T. Hittinger, S. B. Carroll, *Science* **301**, 361–363 (2003).
12. M. Abedin, N. King, *Science* **319**, 946–948 (2008).
13. G. L. Mi  o, M. A. R. Koehl, N. King, R. Stocker, *Limnol. Oceanogr. Lett.* **2**, 37–46 (2017).
14. J. B. Kirkegaard, A. Bouillant, A. O. Marron, K. C. Leptos, R. E. Goldstein, *eLife* **5**, e18109 (2016).
15. R. A. Alegado et al., *eLife* **1**, e00013 (2012).
16. A. Woznica, J. P. Gerdt, R. E. Hulet, J. Clardy, N. King, *Cell* **170**, 1175–1183.e11 (2017).
17. M. Carr et al., *Mol. Phylogenet. Evol.* **107**, 166–178 (2017).
18. C. Nielsen, *Animal Evolution: Interrelationships of the Living Phyla* (Oxford Univ. Press, ed. 3rd, 2012).
19. B. S. C. Leadbeater, *J. Mar. Biol. Assoc. U. K.* **57**, 285–301 (1977).
20. G. J  kely, Evolution of phototaxis. *Philos. Trans. R. Soc. B Biol. Sci.* **364** (2009).
21. M. Watari et al., *J. Biol. Chem.* **294**, 3432–3443 (2019).
22. K. Yoshida, S. P. Tsunoda, L. S. Brown, H. Kandori, *J. Biol. Chem.* **292**, 7531–7541 (2017).
23. L. B. Lamarche et al., *Biochemistry* **56**, 5812–5822 (2017).
24. Y. Tian, S. Gao, S. Yang, G. Nagel, *Biochem. J.* **475**, 1121–1128 (2018).
25. J. L. Spudich, *Trends Microbiol.* **14**, 480–487 (2006).
26. T. Brunet, B. T. Larson, T. A. Linden, N. King, *Choanoecca flexa* transcriptome and predicted nonredundant proteome, Figshare, Version 2 (2019); doi: 10.6084/m9.figshare.8216291.
27. I. Chaves et al., *Annu. Rev. Plant Biol.* **62**, 335–364 (2011).
28. A. Coulon, C. C. Chow, R. H. Singer, D. R. Larson, *Nat. Rev. Genet.* **14**, 572–584 (2013).
29. J. E. P  rez-Ort  n, P. M. Alepuz, J. Moreno, *Trends Genet.* **23**, 250–257 (2007).
30. O. P. Ernst et al., *Chem. Rev.* **114**, 126–163 (2014).
31. D. Arendt, *Int. J. Dev. Biol.* **47**, 563–571 (2003).
32. E. Garc  a-Vald  s, M. G  mila, M. Mulet, J. Lalucat, *Genome Announc.* **6**, e00254–e18 (2018).
33. V. Boswell-Smith, D. Spina, C. P. Page, *Br. J. Pharmacol.* **147** (Suppl 1), S252–S257 (2006).

34. R. E. Weishaar, M. H. Cain, J. A. Bristol, *J. Med. Chem.* **28**, 537–545 (1985).
35. L. A. Wetzel et al., *eLife* **7**, e41482 (2018).
36. A. C. Martin, B. Goldstein, *Development* **141**, 1987–1998 (2014).
37. J. M. Sawyer et al., *Dev. Biol.* **341**, 5–19 (2010).
38. T. A. Richards, T. Cavalier-Smith, *Nature* **436**, 1113–1118 (2005).
39. A. Seb  -Pedr  s, X. Grau-Bov  , T. A. Richards, I. Ruiz-Trillo, *Genome Biol. Evol.* **6**, 290–305 (2014).
40. D. J. Richter, N. King, *Annu. Rev. Genet.* **47**, 509–537 (2013).
41. C. R. Magie, M. Daly, M. Q. Martindale, *Dev. Biol.* **305**, 483–497 (2007).
42. J. Y. Lee, R. M. Harland, *Dev. Biol.* **311**, 40–52 (2007).
43. M. Kov  cs, J. T  th, C. Het  nyi, A. M  n  si-Cs  zmadia, J. R. Sellers, *J. Biol. Chem.* **279**, 35557–35563 (2004).
44. T. Wakatsuki, B. Schwab, N. C. Thompson, E. L. Elson, *J. Cell Sci.* **114**, 1025–1036 (2001).
45. M. Saitoh, T. Ishikawa, S. Matsushima, M. Naka, H. Hidaka, *J. Biol. Chem.* **262**, 7796–7801 (1987).
46. C. P. Heisenberg, Y. Bella  che, *Cell* **153**, 948–962 (2013).
47. D. Gilmour, M. Rembold, M. Leptin, *Nature* **541**, 311–320 (2017).
48. E. N. Marieb, K. Hoehn, *Human Anatomy & Physiology* (Pearson, 2015).
49. M. Nickell, C. Scheer, J. U. Hammel, J. Herzen, F. Beckmann, *J. Exp. Biol.* **214**, 1692–1698 (2011).
50. S. P. Leys, D. Eerkes-Medrano, *Integr. Comp. Biol.* **45**, 342–351 (2005).
51. C. Dayraud et al., *BMC Evol. Biol.* **12**, 107 (2012).
52. K. Pang, M. Q. Martindale, *Dev. Genes Evol.* **218**, 307–319 (2008).
53. S. Armon, M. S. Bull, A. Aranda-D  az, M. Prakash, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E10333–E10341 (2018).
54. P. R. H. Steinmetz et al., *Nature* **487**, 231–234 (2012).
55. T. Brunet et al., *eLife* **5**, e19607 (2016).
56. H. James-Clark, *Mem. Bost. Soc. Nat. Hist.* **1**, 305–340 (1867).
57. W. S. Kent, *Ann. Mag. Nat. Hist.* **5**, 1–17 (1878).
58. S. A. Nichols, B. W. Roberts, D. J. Richter, S. R. Fairclough, N. King, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13046–13051 (2012).
59. D. Arendt, E. Benito-Guti  rrez, T. Brunet, H. Marlow, *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20150286 (2015).
60. W. Franzen, *Zoomorphology* **107**, 349–357 (1988).
61. S. H  hn, A. R. Honerkamp-Smith, P. A. Haas, P. K. Trong, R. E. Goldstein, *Phys. Rev. Lett.* **114**, 178101 (2015).
62. D. J. Dickinson, D. N. Robinson, W. J. Nelson, W. I. Weis, *Dev. Cell* **23**, 533–546 (2012).
63. D. J. Dickinson, W. J. Nelson, W. I. Weis, *BioEssays* **34**, 833–840 (2012).
64. O. Dudin et al., *bioRxiv* 10.1101/563726 (2019).
65. F. Varoquaux et al., *Curr. Biol.* **28**, 3495–3501.e2 (2018).
66. T. Merle, E. Farge, *Curr. Opin. Cell Biol.* **55**, 111–118 (2018); 10.1016/j.cob.2018.07.003.
67. A. Bailles et al., *bioRxiv* 430512 (2019); <https://doi.org/10.1101/430512>.
68. G. O. Mackie, *Q. Rev. Biol.* **45**, 319–332 (1970).
69. X. Zhang, R. H. Cote, *Front. Biosci.* **10**, 1191–1204 (2005).
70. M. Koyanagi et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15576–15580 (2008).
71. D. C. Plachetzki, K. C. Fong, T. H. Oakley, *Proc. Biol. Sci.* **277**, 1963–1969 (2010).
72. G. M. Avelar et al., *Curr. Biol.* **24**, 1234–1240 (2014).
73. D. S. Booth, H. Szmidt-Middleton, N. King, *Mol. Biol. Cell* **29**, 3026–3038 (2018).
74. R. Li, I. Neundorff, F. Nitsche, *bioRxiv* 260190 (2018); <https://doi.org/10.1101/260190>.

ACKNOWLEDGMENTS

We thank the Canadian Institute for Advanced Research for support of field work in Cura  o, P. Keeling for access to the microscope used for Fig. 1C, the staff and students of the 2018 MBL Physiology course, Zeiss for access to an AxioZoom, T. Gerbich and T. Fadero for advice on early phototaxis experiments, D. Jorgens and G. Min (from the UC Berkeley Electron Microscopy Laboratory) for help with SEM, M. Coyle and C. Erikson for advice on transcriptome assembly, M. Sigg and the UC Berkeley Functional Genomic Laboratory for help with iTag sequencing, and D. Bilder, M. Feller, B. Goldstein, and K. Scott for critical feedback on the manuscript. **Funding:** T.A.L. and B.T.L. were supported by NSF GRP Fellowships (B.T.L.: DGE 1106400). T.A.L. was supported by the Berkeley Fellowship for Graduate Study. T.B. was supported by the EMBO long-term fellowship (ALTF 1474-2016) and by the Human Frontier Science Program long-term fellowship (000053/2017-L). **Author contributions:** T.B., B.T.L., and T.A.L.: conceptualization, investigation, methodology, formal analysis, visualization, writing. M.J.A.V.: project administration and resources. K.M.: investigation and visualization. N.K.: conceptualization,

supervision, project administration, and writing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The name-bearing hapantotype of *C. flexa* is deposited with the California Academy of Sciences Invertebrate Zoology Collections (San Francisco) with accession number CASIZ 197968. Raw RNAseq reads used to assemble the *C. flexa* transcriptome have been deposited at the NCBI SRA (accession number/BioProject PRJNA540068, BioSample SAMN11533889). Transcriptome and predicted nonredundant proteome are available on Figshare

(26). mRNA sequences and predicted amino acid sequences for genes of interest are deposited at GenBank: myosin regulatory light chain (MK787241), myosin heavy chain (MK787240), and the four RhoPDEs (MN013138, MN013139, MN013140, and MN013141).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/326/suppl/DC1
Materials and Methods

Figs. S1 to S14
Tables S1 to S4
References (75–94)
Movies S1 to S14

[View/request a protocol for this paper from Bio-protocol.](#)

6 June 2019; accepted 16 September 2019
10.1126/science.aay2346

WATER OXIDATION

An oxyl/oxo mechanism for oxygen-oxygen coupling in PSII revealed by an x-ray free-electron laser

Michihiro Suga^{1,2,*}, Fusamichi Akita^{1,2,*}, Keitaro Yamashita^{3,†}, Yoshiki Nakajima¹, Go Ueno³, Hongjie Li^{1,4}, Takahiro Yamane¹, Kunio Hirata³, Yasufumi Umena¹, Shinichiro Yonekura¹, Long-Jiang Yu¹, Hironori Murakami⁵, Takashi Nomura^{3,4}, Tetsunari Kimura⁶, Minoru Kubo^{3,4}, Seiki Baba⁵, Takashi Kumasaka⁵, Kensuke Tono^{3,5}, Makina Yabashi^{3,5}, Hiroshi Isobe¹, Kizashi Yamaguchi^{7,8}, Masaki Yamamoto³, Hideo Ago^{3,†}, Jian-Ren Shen^{1,†}

Photosynthetic water oxidation is catalyzed by the Mn_4CaO_5 cluster of photosystem II (PSII) with linear progression through five S-state intermediates (S_0 to S_4). To reveal the mechanism of water oxidation, we analyzed structures of PSII in the S_1 , S_2 , and S_3 states by x-ray free-electron laser serial crystallography. No insertion of water was found in S_2 , but flipping of D1 Glu¹⁸⁹ upon transition to S_3 leads to the opening of a water channel and provides a space for incorporation of an additional oxygen ligand, resulting in an open cubane Mn_4CaO_6 cluster with an oxyl/oxo bridge. Structural changes of PSII between the different S states reveal cooperative action of substrate water access, proton release, and dioxygen formation in photosynthetic water oxidation.

Oxygenic photosynthesis by plants, algae, and cyanobacteria converts light energy from the sun into chemical energy in the form of sugar and concurrently releases dioxygen into the atmosphere, thereby sustaining all aerobic life on Earth. The first reaction in oxygenic photosynthesis occurs in PSII, which harbors the oxygen-evolving complex (OEC) that catalyzes stepwise oxidation of water through the S-state cycle (S_i), which includes the ground state S_0 and four oxidized intermediates, S_1 to S_4 (1, 2). Dioxygen is generated in the final step of the S-state transition $S_3 \rightarrow (S_4) \rightarrow S_0$ (3). The OEC contains a Mn_4CaO_5 cluster organized in a distorted-chair form, in which an external Mn is attached to a Mn_3CaO_4 cubane by two μ -oxo O4 and O5 atoms (4, 5). In the OEC, a μ -oxo O5 has unusually long distances from the nearby Mn ions, suggesting weak binding and higher reactivity. O5 may thus serve as one of the substrates for O=O bond formation (1, 2, 4–10). Time-resolved, pump-probe x-ray free electron laser crystallographic analysis of PSII in the S_3 state at 2.35-Å resolution showed an incorporation of an oxygen O6 into the OEC in the vicinity of O5, supporting a dioxygen formation mechanism between O5 and O6 (11). This reaction can proceed via either (i) an oxyl/oxo radical coupling (12), (ii) a nucleophilic attack reaction mechanism (13), or (iii) a peroxide intermediate mechanism (14). Owing to the

limited resolution of the S_3 -state structure, however, the chemical entity of the oxo intermediates (superoxo, peroxy, oxyl/oxo, and hydroxo/oxo) has not been identified, therefore the reaction mechanism was not determined unambiguously.

The OEC is located at a node of five water channels (4) (Fig. 1) involved in proton release, balancing the net charge of the OEC, and inlet of substrate water (1, 15). Dislocation of water W665 in the O4 channel upon progress of the S_i state results in closure of the 15-Å-long, water-mediated hydrogen bonding network (11, 16). We previously attributed this structural change to a proton release through the O4 channel, whereas others have argued that W665 serves as the source of the O6 atom involved in O=O bond formation (17, 18). Lack of a high-resolution structure of the S_2 state has contributed to uncertainty about the structural changes leading to or from this state. To address these issues, we fixed PSII in the S_1 , S_2 , and S_3 states, and the triply flashed (3F) state by a cryo-trapping method with microcrystals of PSII from the thermophilic cyanobacterium *Thermosynechococcus vulcanus*. Structures at 2.15- to 2.50-Å resolution using fixed-target serial femtosecond crystallography (SFX) are consistent with the structures at room temperature determined previously (11, 16) but allow more accurate determination of interatomic distances in the OEC, revealing the chemical entity of the intermediate O5 and O6 species and the corresponding changes in the protein-ligand environment

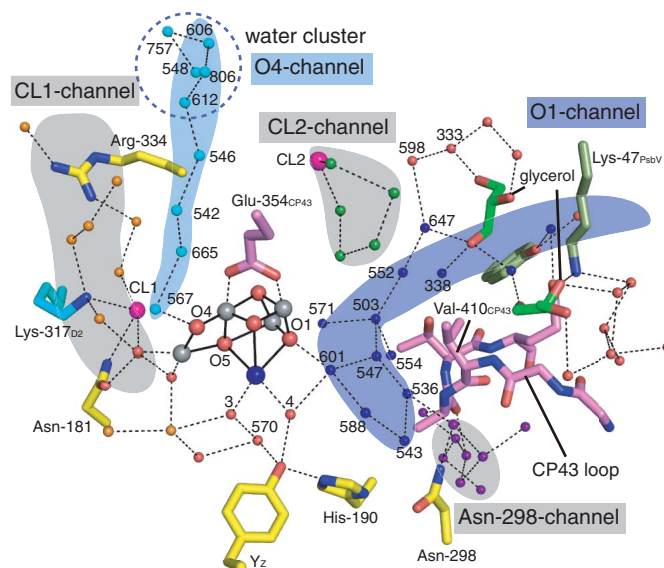


Fig. 1. Water networks in PSII.

Five hydrogen-bonded water networks surround the OEC (PDB ID 4UB6) (5). The upper-bound distance for the hydrogen bonds shown is 3.3 Å. CL1 and CL2 represent chloride ions found near the OEC.

¹Research Institute for Interdisciplinary Science and Graduate School of Natural Science and Technology, Okayama University, 3-1-1 Tsushima Naka, Okayama 700-8530, Japan. ²Japan Science and Technology Agency, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. ³RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan. ⁴Department of Picobiology, Graduate School of Life Science, University of Hyogo, 3-2-1 Kouto, Kamigori-cho, Ako-gun, Hyogo 678-1297, Japan. ⁵Japan Synchrotron Radiation Research Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5198, Japan. ⁶Department of Chemistry, Graduate School of Science, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan. ⁷The Institute for Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan. ⁸Institute for Nanoscience Design, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan.

*These authors contributed equally to this work.

†Corresponding author. Email: michisuga@okayama-u.ac.jp (M.S.); ago@spring8.or.jp (H.A.); shen@cc.okayama-u.ac.jp (J.-R.S.) ‡Present address: Department of Biological Science, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

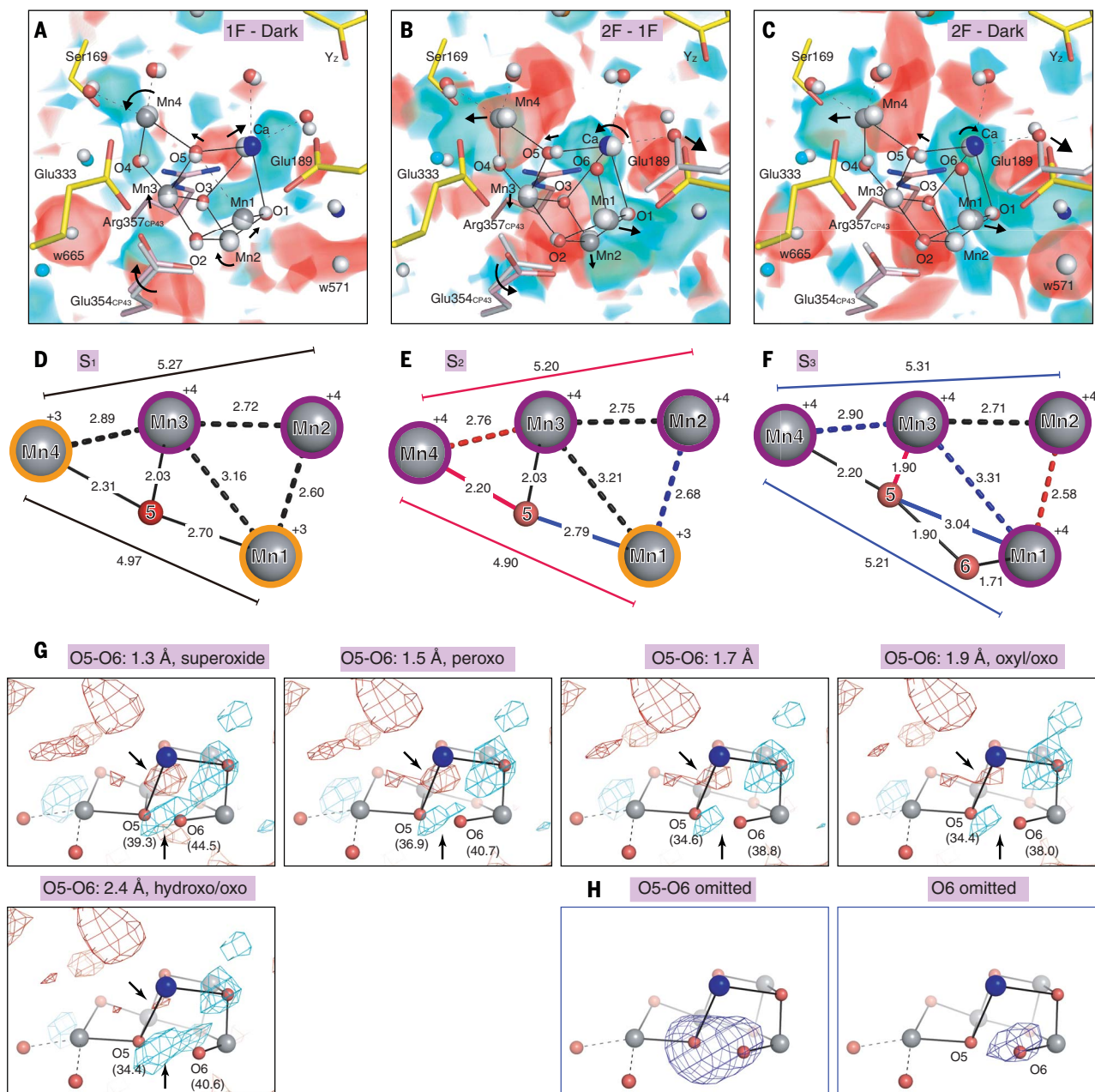


Fig. 2. Structural changes of the OEC during the S_i -state transitions.

(A to C) OEC structures superimposed with $F_{\text{obs}} - F_{\text{calc}}$ isomorphous difference Fourier maps of (A) 1F minus dark, (B) 2F minus 1F, and (C) 2F minus dark datasets. Structures before and after S_i -state transition are shown in gray and color, respectively. Difference maps are contoured at -3σ (red) and 3σ (cyan). Structural changes consistent with isomorphous difference Fourier maps are represented by black arrows, where larger arrowheads represent the larger structural changes. (D to F) Interatomic distances (angstroms) of the OECs in the (D) S_1 , (E) S_2 , and (F) S_3 states. Blue and red lines indicate elongation and shortening of the interatomic distances compared with the structure in the

precedent S_i state. Presumed Mn (+4) and Mn (+3) cations are shown.

(G) The $F_{\text{obs}} - F_{\text{calc}}$ difference Fourier maps contoured at -2.2σ (red) and $+2.2\sigma$ (cyan) after structural refinement by fixing the distance between O5 and O6 at 1.3 Å (superoxide), 1.5 Å (peroxo), 1.7 Å, 1.9 Å (oxyl/oxo), and 2.4 Å (hydroxo/oxo), respectively. Values in parentheses are temperature factors (square angstroms) of O5 and O6. Black arrows indicate the residual electron densities that were affected by the interatomic distances between O5 and O6.

(H) The $F_{\text{obs}} - F_{\text{calc}}$ difference Fourier maps contoured at 5σ (blue) when both O5 and O6, or just O6, were omitted. Colors used here are the same in all figures unless otherwise noted.

responsible for the proton exit and water inlet during the S-state cycle.

Structural determination

Single-shot diffraction images were collected in a fixed-target data collection manner at a

cryogenic temperature (5, 19), in which PSII microcrystals were used (fig. S1, see supplementary materials and methods). Compared with SFX of PSII using a grease matrix as the injection medium at room temperature (17), this method reduced the sample consump-

tion by one order of magnitude and ensured low background images, allowing us to collect diffraction images with high hit and index rates, which yielded 2.15-Å resolution datasets for PSII in the S_1 , S_2 , and S_3 states (dataset 1 in tables S1 and S2). We collected datasets of a 3F

state and S_1 , S_2 , and S_3 states (dataset 2) at 2.35- to 2.50-Å resolution independently to confirm the reproducibility of the light-induced structural changes and to examine the structural changes beyond the S_3 state.

Strong peaks in the isomorphous difference Fourier maps calculated between each state (Fig. 2, A to C, and fig. S3) indicate successful detection of the structural changes induced by the flash illuminations. Most peaks were localized in the vicinity of OEC, in agreement with the previous observations at room temperature by the SFX method (11) (Figs. 2 to 4 and fig. S3). However, structural changes in the quinone Q_B -binding site were diminished substantially, reflecting the relaxed protein environment due to the longer delay time after flash illumination in the present study (1 s in

the fixed-target method versus 10 ms in the SFX method). Pairs of positive and negative density peaks were more clearly visible in the current difference maps around the OEC owing to the high resolution achieved as well as the high isomorphism between the different S-state datasets. Changes in the interatomic distances between the four Mn atoms accompanying the S_2 -state transition (table S4) reflect changes of the oxidation states of the Mn ions. Difference maps between sequential states allowed us to define the order of serial structural changes induced by S-state progression.

The S_1 and S_2 states

The Mn-Mn distances in the S_1 state in our structure (Fig. 2D and table S4) are similar to previous results (5, 11, 16, 20), with the ex-

ception being that the shortest distance (Mn1-Mn2, 2.60 Å) is slightly shorter than what was previously reported. The S_2 state may adopt either an open or a closed cubane form according to theoretical studies (21, 22). The open cubane corresponds to the structure giving rise to the $S = 1/2$, $g = 2.0$ electron paramagnetic resonance (EPR) multiline signal (S , spin; g , g-factor, a dimensionless quantity that characterizes the magnetic moment and angular momentum of an atom), and the closed cubane to the $S = 5/2$, $g = 4.1$ signal (21, 22). The S_2 state in our structure is in the open cubane form, in agreement with the absence of the $g = 4.1$ EPR signal in cyanobacterial PSII under normal conditions (23). Difference density analysis reveals that, upon S_1 -to- S_2 transition, Mn4 is shifted toward Glu³³³ and Ca is moved toward Yz (Tyr¹⁶¹) slightly (Fig. 2A). There is a shift of O5 toward Mn4 and a weakening or breakage of the oxy bridge between O5 and Mn1 in the S_2 state (fig. S4). These changes are consistent with Mn4 oxidation during the S_1 -to- S_2 transition, giving rise to a bipyramidal five coordinated Mn1 and an open cubane OEC structure consistent with a charge distribution of (Mn1, Mn2, Mn3, Mn4) = (III, IV, IV, IV). Slight changes in Mn-Mn distances are observed in the S_2 state (Fig. 2E), resulting in an increased homogeneity in these distances consistent with x-ray absorption fine structure measurements (24).

In addition to the structural changes of OEC, strong negative densities were found for two water molecules, W665 and W571, that are hydrogen-bonded with ligand residues of the OEC (Fig. 2, A and C). These changes have been found in the S_1 -to- S_3 transition (11), but the present results show that these two water molecules already become highly disordered in the S_1 -to- S_2 transition; therefore, the possibility of W665 as the substrate for O=O bond formation (17, 18) is unlikely. W665 is the second water molecule from O4 within the 15-Å-long O4 water chain that ends at a water cluster consisting of five water molecules exposed to the luminal solution (Fig. 1). This water cluster may accept a proton released from OEC as a protonated Eigen cation through the Grotthuss-type proton transfer (25), consistent with Fourier transform infrared spectroscopy, which showed changes in a highly polarized hydrogen-bond network (26–28) or possible formation of a $n\text{H}_2\text{O}(\text{H}_3\text{O})^+$ cluster in the S_2 state (29, 30). Despite a long distance between the water cluster and the OEC, change was observed in both electron density and the shape of the water cluster (Fig. 3, A and B). These changes may result from positive charge accumulated during the S_1 -to- S_2 transition, possibly by the ejection of a proton from the OEC through W665. The increased mobility of W665 may be required

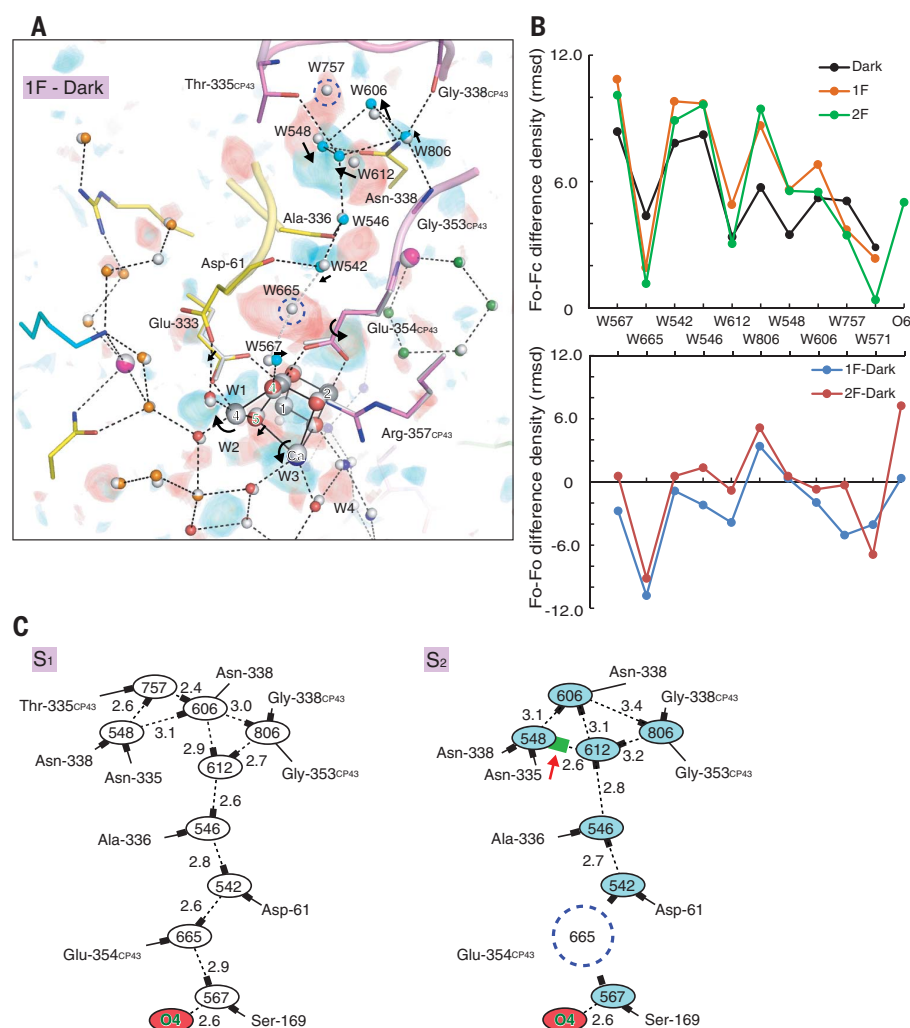


Fig. 3. Structural changes in the O4 channel. (A) $F_{\text{obs}} - F_{\text{calc}}$ isomorphous difference Fourier map of 1F minus dark contoured at 3σ superimposed with the hydrogen-bonding network in the O4 channel. Dislodged water molecules are highlighted by blue circles. (B) $F_{\text{obs}} - F_{\text{calc}}$ when water or O6 were omitted (top) or $F_{\text{obs}} - F_{\text{obs}}$ (bottom) difference density values at water sites in the O4 channel. rmsd, root mean square deviation. (C) Possible proton transfer mechanism in the O4 channel. Hydrogen-bonding networks before (left) and after (right) proton transfer. The putative proton stored at the water cluster is depicted in green and indicated by a red arrow.

to prevent the backflow of the proton by disconnecting the water chain (Fig. 3C) and is accompanied by a slight shift of the side chain of CP43 Glu³⁵⁴ hydrogen-bonded with W665. W571 also increases its mobility on the S₂-to-S₃ transition (Figs. 2, A and C, and 3B), likely as a result of the accumulated positive charges on the OEC. W665 appeared in its position again in the transition of S₃ to 3F state (supplementary text, fig. S3K), indicating regeneration of the long O4 hydrogen bond network after the S₃ state for proton transfer in the next reaction cycle.

The oxyl/oxo species in the S_3 state

Larger difference densities were observed around the OEC for the S_2 -to- S_3 transition than for the S_1 -to- S_2 transition (Fig. 2, A and B, and fig. S3, B and C). Mn1, Mn2, and Mn3 are static, but Mn4 is moved toward Ser¹⁶⁹ (Fig. 2B), resulting in an increase in the Mn1-Mn4 distance (Fig. 2, E and F). Flipping of the side chain of Glu¹⁸⁹, also seen at room temperature (*II*), provides an open space in the vicinity of O5, enabling the insertion of O6 (Fig. 4, A and B). The improved resolution of the dataset allowed us to identify the position of O6 clearly in both $F_{\text{obs}} - F_{\text{obs}}$ and $F_{\text{obs}} - F_{\text{calc}}$ difference maps and thus determine its chemical structure unambiguously (Fig. 2, B, C, and H). By altering the O5-O6 distance and examining the residual densities in the $F_{\text{obs}} - F_{\text{calc}}$ difference Fourier map, we found that a distance of 1.9 Å resulted in the weakest residual densities (Fig. 2G) (see materials and methods). This distance is slightly longer than the 1.5 Å we reported previously (*II*) and consistent with an oxyl/oxo pair for the O5 and O6 species. Distances for superoxo (1.3 Å), peroxo (1.5 Å), and hydroxo/oxo (2.4 Å) species can be excluded on the basis of their increased residual densities. Furthermore, a hydroxo/oxo pair cannot be accommodated by the current structure owing to the limited space available, unless the interatomic distance between O5 and Mn4 is shorter by 0.3 Å. A peroxo species fits with the electron density similarly to that of the oxyl/oxo species, explaining why we could not discriminate in the previous study (*II*); however, the oxyl/oxo species resulted in less residual electron densities in the $F_{\text{obs}} - F_{\text{calc}}$ map and evenly distributed temperature factors for O5 and O6. The interatomic distances of OEC in the S_3 state agree well with those of the theoretically optimized structure of S_3 when the oxyl/oxo species was assumed in the open cubane form (table S5) (*3I*). Thus, we conclude that the O5 and O6 pair is in an oxyl/oxo form, which would be consistent with an oxyl/oxo coupling mechanism for the O=O bond formation (*3I*).

Movement of D1 Glu¹⁸⁹, the only monodentate carboxylate ligand of the Mn₄CaO₅ cluster, by 0.5 Å during the S₂-to-S₃ transition

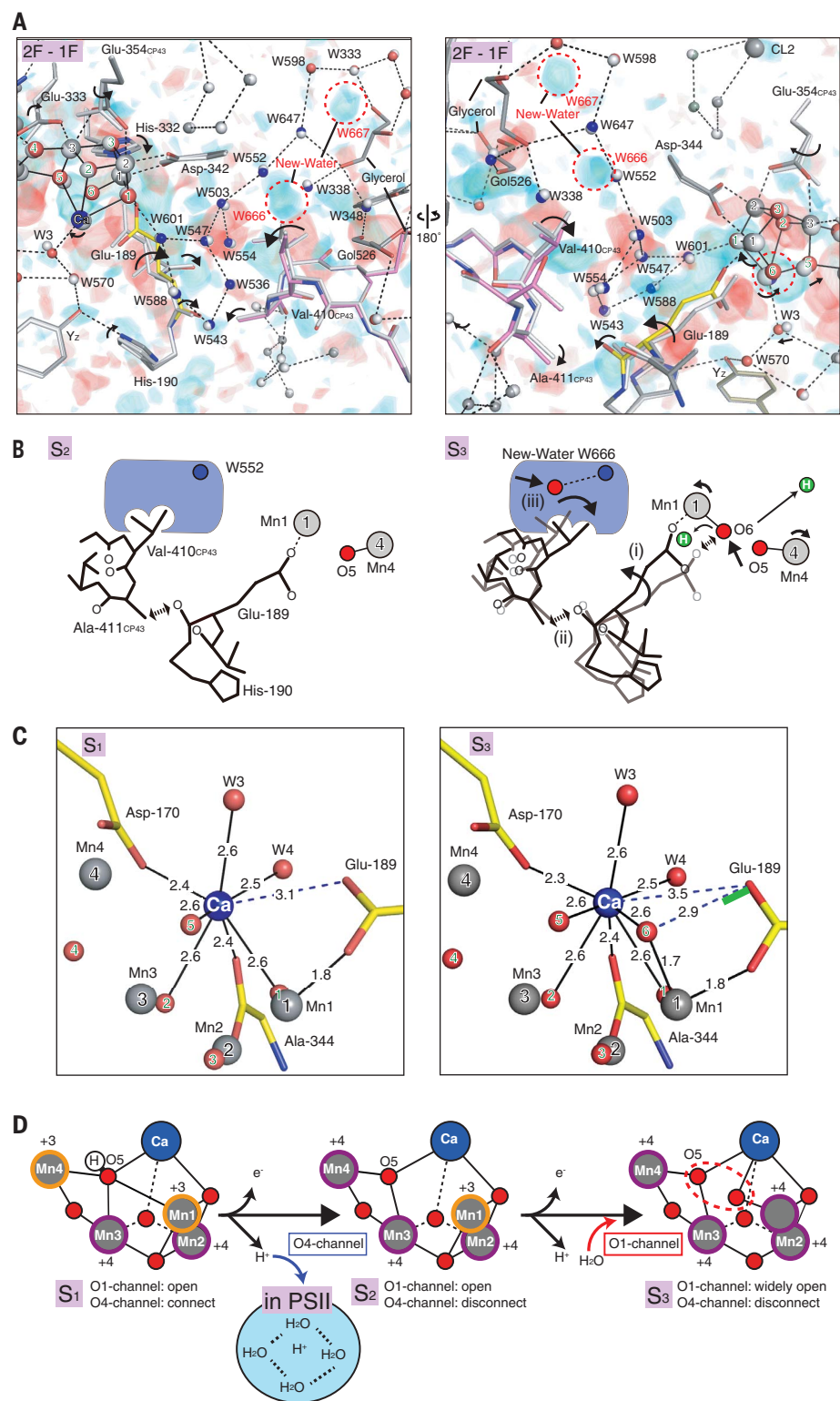


Fig. 4. Structural changes in the water-inlet O1 channel and around the Ca^{2+} ion. (A) Isomorphous difference map (3σ) superimposed with the O1 channel. O6 and two water molecules that appear in the S_3 state are marked with a red dashed circle. (B) Model of structural changes in the O1 channel. (C) The ligand environment of Ca^{2+} in the S_1 and S_3 states. Weak interactions between Glu¹⁸⁹ and Ca^{2+} , and between Glu¹⁸⁹ and O6, are shown as dashed lines. Possible hydrogen atom transferred from O6 to Glu¹⁸⁹ is shown as a solid green line. (D) Schematic structures of OEC and the states of the O1 and O4 channels in the S_1 – S_3 states. The water molecule from the O1 channel may be incorporated into O6 in S_3 or refill the water molecules waiting for the next catalytic reaction.

causes an elongation of the distance between its OE1 and Ca^{2+} from 3.1 Å in the S_1 state to 3.5 Å in the S_3 state (Fig. 4C). The Ca^{2+} ion can be considered to have a coordination number of 7.5 or pseudo-eight-coordinated in the S_1 state, because it has seven “normal” ligands plus D1 Glu¹⁸⁹. In the S_3 state, the interaction between D1 Glu¹⁸⁹ and Ca^{2+} is much weaker, but the newly inserted O6 provides an eighth ligand to Ca^{2+} with a distance of 2.6 Å (Fig. 4C), which is made possible by the ability of Ca^{2+} to adopt coordination numbers from four to eight (32). The two hydrogen atoms brought by O6 are, respectively, ejected from the active site during the S_2 -to- S_3 transition and accepted by D1 Glu¹⁸⁹, which forms a hydrogen bond with O6 (Fig. 4C), giving rise to the oxyl/oxo species between O6 and O5.

The mechanism for substrate water delivery

Flipping of the Glu¹⁸⁹ side chain is correlated with motions in a short loop of CP43, including CP43 Val⁴¹⁰, that restrict the size of the O1 channel (Fig. 4, A and B). Paired positive and negative densities were found around CP43 Val⁴¹⁰ indicating a movement of this residue toward Glu¹⁸⁹ by 0.5 Å in the S_3 state. This movement substantially widens the channel radius (Fig. 4, A and B), which may allow water molecules to come in during the S-state transition. Strong negative difference densities were found in positions overlapping with W503, W547, and W554 as well as glycerol-526 (Gol526) present in the channel, reflecting increased mobility of these molecules in response to the channel opening in the S_3 state. Two positive difference densities in the hydrogen-bonded water network indicate the insertion of water molecules W666 and W667 into the O1 channel (Fig. 4, A and B). The O1 channel likely serves as a conduit for substrate water entry into the OEC, and the residues around CP43 Val⁴¹⁰ may serve as a “valve” to control the water-inlet channel. Movements in this region coincide with incorporation of O6 into the open cubane structure of OEC during the S_2 -to- S_3 transition. Thus, Glu¹⁸⁹ plays a pivotal role in coupling oxidation of the OEC with the opening of the water channel and delivery of the substrate into the OEC.

Outlook

The structural changes related to proton release, water inlet, and O=O bond formation during the S-state cycle are summarized in Fig. 4D. The displacement of W665 in the O4

channel suggests a proton transfer from the OEC upon S_1 -to- S_2 transition. Considering that theoretical calculations favor O5 as a hydroxide ion in S_1 and an oxide ion in the open cubane S_2 structure (33), and that there is no proton release to the bulk solution in the S_1 -to- S_2 transition, a proton is likely ejected from the O5 site and stored as a protonated Eigen cation in PSII. The structural changes observed in the O4 channel thus provide insights into the timing of proton release and an elegant way to prevent the backflow of the released proton. Insertion of O6 occurs in the S_2 -to- S_3 transition, providing an oxyl/oxo species in the S_3 state. Structural changes of the Glu¹⁸⁹ side chain serve to couple oxidation of the catalytic site with substrate water access, proton release, and O=O bond formation via the oxyl/oxo coupling mechanism. The structure and O=O bond formation mechanism revealed here should serve as an important blueprint for rational design of artificial catalysts that have a capability for oxidizing water by visible light.

REFERENCES AND NOTES

1. J. R. Shen, *Annu. Rev. Plant Biol.* **66**, 23–48 (2015).
2. N. Cox, D. A. Pantazis, F. Neese, W. Lubitz, *Acc. Chem. Res.* **46**, 1588–1596 (2013).
3. B. Kok, B. Forbush, M. McGloin, *Photochem. Photobiol.* **11**, 457–475 (1970).
4. Y. Umena, K. Kawakami, J. R. Shen, N. Kamiya, *Nature* **473**, 55–60 (2011).
5. M. Suga et al., *Nature* **517**, 99–103 (2015).
6. S. Yamanaka et al., *Chem. Phys. Lett.* **511**, 138–145 (2011).
7. P. E. Siegbahn, *Biochim. Biophys. Acta* **1827**, 1003–1019 (2013).
8. P. E. Siegbahn, *Chemistry* **14**, 8290–8302 (2008).
9. N. Cox et al., *Inorg. Chem.* **50**, 8238–8251 (2011).
10. I. Zaharieva et al., *Biochemistry* **55**, 4197–4211 (2016).
11. M. Suga et al., *Nature* **543**, 131–135 (2017).
12. X. Li, P. E. Siegbahn, *Phys. Chem. Chem. Phys.* **17**, 12168–12174 (2015).
13. E. M. Sproviero, J. A. Gascón, J. P. McEvoy, G. W. Brudvig, V. S. Batista, *J. Am. Chem. Soc.* **130**, 3428–3442 (2008).
14. H. Isobe, M. Shoji, J. R. Shen, K. Yamaguchi, *Inorg. Chem.* **55**, 502–511 (2016).
15. H. Dau, M. Haumann, *Coord. Chem. Rev.* **252**, 273–295 (2008).
16. J. Kern et al., *Nature* **563**, 421–425 (2018).
17. J. Wang, M. Askerka, G. W. Brudvig, V. S. Batista, *ACS Energy Lett.* **2**, 2299–2306 (2017).
18. K. Kawashima, T. Takaoka, H. Kimura, K. Saito, H. Ishikita, *Nat. Commun.* **9**, 1247 (2018).
19. K. Hirata et al., *Nat. Methods* **11**, 734–736 (2014).
20. A. Tanaka, Y. Fukushima, N. Kamiya, *J. Am. Chem. Soc.* **139**, 1718–1721 (2017).
21. H. Isobe et al., *Dalton Trans.* **41**, 13727–13740 (2012).
22. D. A. Pantazis, W. Ames, N. Cox, W. Lubitz, F. Neese, *Angew. Chem.* **51**, 9935–9940 (2012).
23. A. Bousset, H. Kuhl, S. Un, M. Rögner, A. W. Rutherford, *Biochemistry* **37**, 8995–9000 (1998).

24. C. Glöckner et al., *J. Biol. Chem.* **288**, 22607–22620 (2013).
25. K. Saito, A. W. Rutherford, H. Ishikita, *Nat. Commun.* **6**, 8488 (2015).
26. T. Noguchi, M. Sugiura, *Biochemistry* **39**, 10943–10949 (2000).
27. T. Noguchi, H. Suzuki, M. Tsuno, M. Sugiura, C. Kato, *Biochemistry* **51**, 3205–3214 (2012).
28. R. J. Debus, *Biochemistry* **53**, 2941–2955 (2014).
29. J. M. Headrick et al., *Science* **308**, 1765–1769 (2005).
30. B. C. Polander, B. A. Barry, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10634–10639 (2013).
31. H. Isobe, M. Shoji, T. Suzuki, J. R. Shen, K. Yamaguchi, *J. Chem. Theory Comput.* **15**, 2375–2391 (2019).
32. I. Dokmanić, M. Sikić, S. Tomić, *Acta Crystallogr. D* **64**, 257–263 (2008).
33. M. Shoji et al., *Chem. Phys. Lett.* **623**, 1–7 (2015).

ACKNOWLEDGMENTS

The x-ray free electron laser experiments were performed at SACL with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (proposal nos. 2016A8051, 2016B8070, 2017A8040, 2017B8015, 2017B8028, 2018A8009, and 2018A8010), and we thank the staff at SACL for their help. The best condition for diffraction data collection was determined at beamlines 41XU and 44XU in SPring-8 (proposal nos. 2016A2542, 2016A6621, 2016B2542, 2016B6621, 2017A2535, 2017A6724, 2017B2535, 2017B6724, 2018A2530, 2018A6822, 2018B2530, and 2018B6822). **Funding:** This research was supported by the Platform Project for Supporting Drug Discovery and Life Science Research [Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)] from AMED under grant JP18am0101070. This research was also supported by KAKENHI grants JP16H06162, JP16H06296, and JP17H05884 (M.S.); JP17H06434 (J.-R.S.); and JP16K21181 (F.A.). Support was also provided by JST, PREST grants JPMJPR18G8 (M.S.) and JPMJPR16P1 (F.A.), and RIKEN Pioneer Project “Dynamic Structural Biology” to M.K. and M.Yam. **Author contributions:** J.-R.S., M.S., M.Yam., and H.A. conceived the project. G.U., H.M., S.B., T.Ku., H.A., and M.Yam. developed the data collection setup. M.K., T.Ki., and T.N. set up a laser system for the sample preparation. K.T. and M.Yab. contributed to beamline operation. F.A. and Y.N. prepared the samples and grew the crystals. M.S., H.L., T.Y., Y.U., S.Y., L.-J.Y., and J.-R.S. collected the diffraction data. K.Yamas., H.L., T.Y., and M.S. processed the diffraction data. M.S. analyzed the structure. H.I. and K.Yamag. discussed the atomic coordinates. M.S. and J.-R.S. wrote the manuscript. All authors contributed to the discussion and improvement of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the following IDs: 6JLJ for S_1 (dataset 1), 6JLK for S_2 (dataset 1), 6JLL for the S_3 (dataset 1), 6JLM for S_1 (dataset 2), 6JLN for S_2 (dataset 2), 6JLO for S_3 (dataset 2), and 6JLP for 3F. All other data used in this study are presented in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/334/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
Tables S1 to S6
References (34–44)

[View/request a protocol for this paper from Bio-protocol.](#)

15 April 2019; accepted 9 September 2019
10.1126/science.aax6998

BIODIVERSITY

The geography of biodiversity change in marine and terrestrial assemblages

Shane A. Blowes^{1,2*}†, Sarah R. Supp^{3*}†, Laura H. Antão^{4,5,6}, Amanda Bates⁷, Helge Bruelheide^{1,8}, Jonathan M. Chase^{1,2}, Faye Moyes⁴, Anne Magurran⁴, Brian McGill⁹, Isla H. Myers-Smith¹⁰, Marten Winter¹, Anne D. Bjorkman¹¹, Diana E. Bowler¹, Jarrett E. K. Byrnes¹², Andrew Gonzalez¹³, Jes Hines^{1,14}, Forest Isbell¹⁵, Holly P. Jones¹⁶, Laetitia M. Navarro^{1,8}, Patrick L. Thompson¹⁷, Mark Vellend¹⁸, Conor Waldock¹⁹, Maria Dornelas^{4*}

Human activities are fundamentally altering biodiversity. Projections of declines at the global scale are contrasted by highly variable trends at local scales, suggesting that biodiversity change may be spatially structured. Here, we examined spatial variation in species richness and composition change using more than 50,000 biodiversity time series from 239 studies and found clear geographic variation in biodiversity change. Rapid compositional change is prevalent, with marine biomes exceeding and terrestrial biomes trailing the overall trend. Assemblage richness is not changing on average, although locations exhibiting increasing and decreasing trends of up to about 20% per year were found in some marine studies. At local scales, widespread compositional reorganization is most often decoupled from richness change, and biodiversity change is strongest and most variable in the oceans.

Humans are reshaping biodiversity patterns. Against a background of elevated extinction rates (1, 2), local biodiversity change results from multiple interacting drivers that influence the abundance and distribution of species. Different regions of the globe are projected to experience different trends in biodiversity change, particularly those caused by variations in the strength of drivers such as land-use intensity (3) and climate change (4). There are widespread changes in the identities of species that live in any one location (species composition), whereas shifts in the numbers of species (species richness) show mixed patterns, with increasing, decreasing, or static trends (5–9). However, the spatial distribution of the locations most affected is unknown. Here, we map biodiversity change, in terms of species richness and composition, to uncover the geography of biodiversity change. Our analysis compares assemblage time series across the marine, terrestrial, and freshwater realms, different biomes, and latitudinal bands (i.e., polar, temperate, and tropical).

Both biodiversity and its change are unevenly distributed on the planet (10, 11) and unevenly sampled (12–15). Species densities typically decline drastically from the tropics to the poles, and the identities of species differ

across continents and oceans. Hence, knowing which locations are undergoing different types of net change in biodiversity is critical to understanding how biodiversity is changing globally. Detecting geographic variation in biodiversity trends will not only improve our understanding of how global biodiversity is changing but will also inform conservation prioritization. Specifically, by identifying the regions of the planet that are changing more, we will be better placed to make informed decisions about the spatial distribution of biodiversity vulnerability and about where to prioritize reactive (such as restoration) and proactive (protection) conservation actions (16). In addition, quantifying this spatial distribution will refine hypotheses about the drivers of biodiversity change.

Spatial patterns in biodiversity change are the combined result of species changing their distributions, entering and leaving local communities, going extinct, or speciating. These processes are affected by many drivers, which themselves are spatially heterogeneous (17, 18) and differ between the marine and terrestrial realms (19). For example, spatial overlap between climate change and other drivers of change is greater in the marine realm than in the terrestrial realm (19). Moreover, species sensitivities to climate change (i.e., temperature

increases) are also greater in the marine realm (20, 21). When coupled with ecological differences between realms, such as fewer barriers to dispersal and greater colonization rates in marine ecosystems (22, 23), these differences may result in greater compositional change in marine compared with terrestrial assemblages (21). Therefore, we predict that biodiversity in the marine realm has changed more than it has in the terrestrial realm. Changes in community composition are not necessarily associated with changes in species richness if species gains and losses are approximately balanced (5–9, 24, 25). However, in regions where land-use intensity is high (26) or where range sizes contract in response to climate change (27), species extirpations could result in a decrease in species richness. Conversely, in regions subject to high rates of species introductions (28–31) or high connectivity, or where ranges expand (32, 33) or species are broadly favored by land-use change (34), species richness could increase. Hence, we expect there to be variation in biodiversity change across different biomes and geographical regions of the planet. There may also be latitudinal differences in biodiversity change. For example, tropical regions are entering climatic conditions with no present-day equivalents (35), and intensification of land-use change is more recent than in temperate regions (36); therefore, richness loss may be more prevalent and more extreme in tropical latitudes. Here, we investigated whether biodiversity change differs in magnitude between the realms and if strong geographic patterns exist in the changes to species richness and composition across realms, latitudes, or regions.

We examined geographic variation in patterns of change in both species richness and composition using local assemblage time series from across the globe (37, 38) (fig. S1). The BioTIME database is currently the largest compilation of assemblage time series, and our analysis included 239 independent studies (table S1). Because spatial extent varied considerably among studies, we used a gridding method (96-km² hexagonal cells) (39) to partition the 126 studies that had multiple sampling locations and large spatial extents (38); 113 studies were not partitioned because they were contained within a single grid cell. This resulted in 51,932 individual

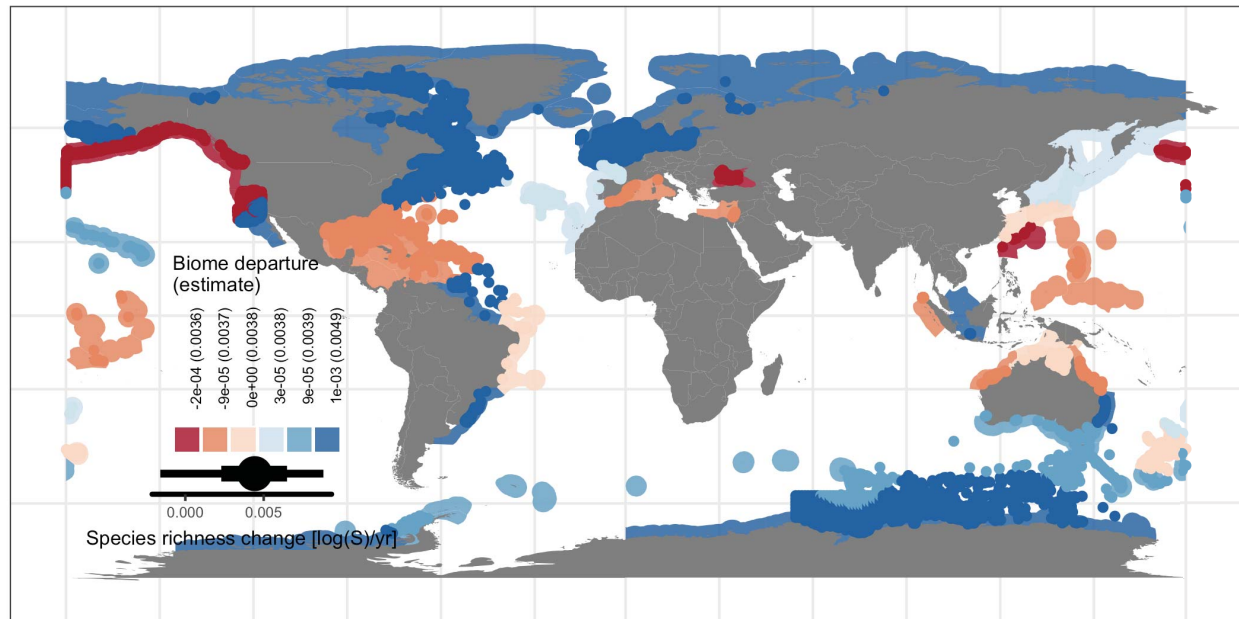
¹German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany. ²Martin Luther University Halle-Wittenberg, Institute of Computer Science, Halle (Saale), Germany. ³Data Analytics Program, Denison University, Granville, OH, USA. ⁴Centre for Biological Diversity, School of Biology, University of St. Andrews, St. Andrews, UK. ⁵Department of Biology and CESAM, Universidade de Aveiro, Aveiro, Portugal. ⁶Research Centre for Ecological Change, Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland. ⁷Department of Ocean Sciences, Memorial University of Newfoundland, Newfoundland, Canada. ⁸Martin Luther University Halle-Wittenberg, Institute of Biology/Geobotany and Botanical Garden, Halle (Saale), Germany. ⁹School of Biology and Ecology, University of Maine, Orono, ME, USA. ¹⁰School of GeoSciences, University of Edinburgh, Edinburgh, UK. ¹¹Senckenberg Gesellschaft für Naturforschung, Biodiversity and Climate Research Centre (BiK-F), Frankfurt am Main, Germany. ¹²Department of Biology, University of Massachusetts Boston, Boston, MA, USA. ¹³Department of Biology, Quebec Centre for Biodiversity Science, McGill University, Montreal, QC, Canada. ¹⁴Leipzig University, Institute of Biology, Leipzig, Germany. ¹⁵Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN, USA. ¹⁶Department of Biological Sciences and Institute for the Study of the Environment, Sustainability, and Energy, Northern Illinois University, DeKalb, IL, USA. ¹⁷Department of Zoology, University of British Columbia, Vancouver, BC, Canada. ¹⁸Département de biologie, Université de Sherbrooke, Sherbrooke, QC, Canada. ¹⁹Ocean and Earth Science, National Oceanography Centre, University of Southampton, Southampton, UK & Life Sciences, Natural History Museum, Cromwell Road, London, UK. *Corresponding author. Email: sablowes@gmail.com (S.A.B.); supps@denison.edu (S.R.S.); maadd@st-andrews.ac.uk (M.D.) †These authors contributed equally to this work.

local assemblage time series, with each time series composed of samples from only one study. This means that important study-level considerations (e.g., sampling method) were consistent within each time series. After fur-

ther filtering by sampling completeness and standardization by sample-based rarefaction (38), these time series became the lowest level in our hierarchical models of temporal trends. Temporal extent and start date vary substan-

tially within these data: time series span from the late 1800s to the present, though most data come from the past 40 years (fig. S2), and we examined the sensitivity of our results to this heterogeneity.

A Marine



B Terrestrial and freshwater

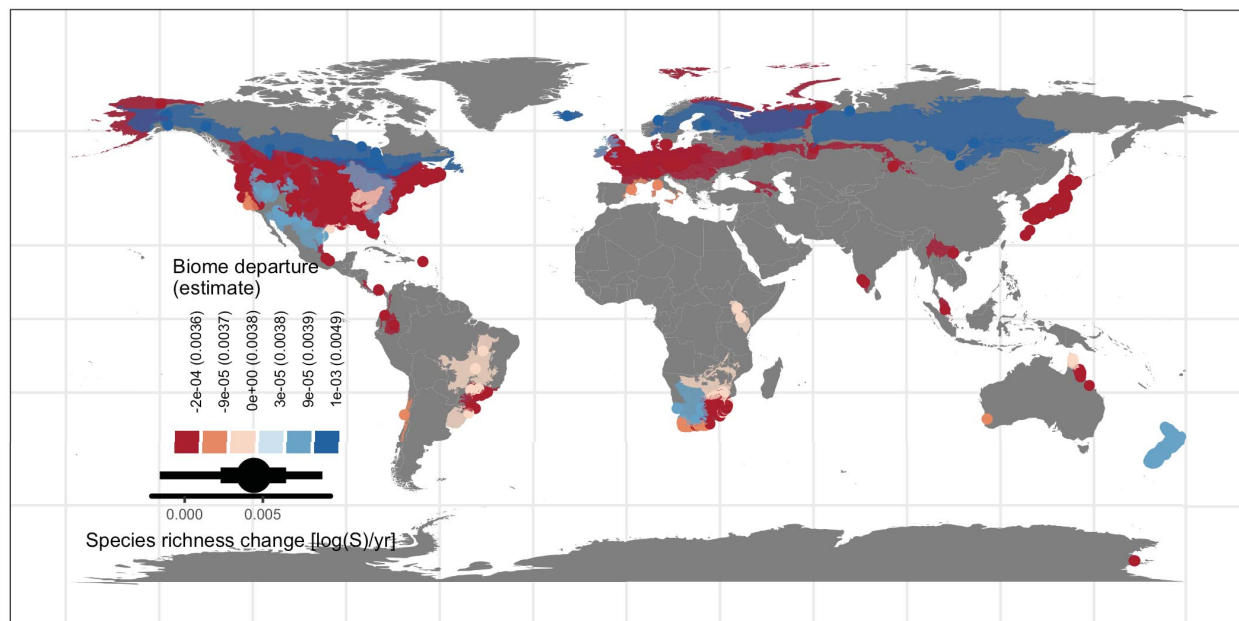


Fig. 1. Species richness change maps showing departures from the overall trend for marine and terrestrial biomes. Inset shows the overall trend in assemblage species richness change; bar depicts 50% (thick) and 90% (thin) credible intervals. Shading on the map represents positive (blue; faster increases in species richness than average) and negative (red; slower increases in species richness than average) departures for each biome from the overall average species richness change (0.004 log species per year). Numbers in the

inset denote the departure and the biome-level (overall + departure) estimate in parentheses. The 90% credible intervals for all biome-level estimates overlap zero. **(A)** Marine biomes ($n = 33$) show both positive and negative departures from the overall trend, with more negative departures in the tropics, whereas there are no latitudinal trends in **(B)** terrestrial ($n = 10$) and freshwater ($n = 5$) biomes, which also show both positive and negative departures from the overall trend.

Biodiversity trends across the globe

To examine geographic patterns in biodiversity change, we quantified realm, latitudinal, and regional departures from the overall trends of richness and composition change using hierarchical generalized linear models. We first nested the cell-level time series within the 239 original studies to control for effects of sampling methods and nonindependence of cell-level time series that came from a single study. Throughout our analysis, we controlled for not having the same taxa sampled everywhere by including taxon in our models. Our first model, the biome-taxon model, nested studies into nine taxonomic-habitat groupings that were further nested within 48 biomes [defined by the Ecoregions of the World datasets available from *The Nature Conservancy* website; http://maps.tnc.org/gis_data.html, TNC terrestrial regions dataset (40–43)]; this resulted in 321 biome-taxon study combinations grouped within 105 specific biome-taxon combinations. The 48 biomes, including 33 marine (41), 10 terrestrial (44), and 5 freshwater (42), represent geographic regions of the world and allowed us to characterize spatial patterns as biome-level departures from the overall trend of biodiversity change for each realm. The taxonomic-habitat groupings, dictated by specifications in the original studies, were amphibians, benthos, birds, fish, invertebrates, mammals, marine invertebrates and plants, plants, and multiple taxa, and were included to contrast and control for differences in trends among taxa within the different biomes. We examined the robustness of our biome-taxon models, and the spatial patterns they identified, by fitting two complementary hierarchical models with simpler geographic structures. Here, we focus on the simplest model (referred to as the realm-latitude-taxon model) and present a model of intermediate complexity in the supplement (38). All models grouped cells within studies at the lowest level, and the realm-latitude-taxon model grouped studies into 29 specific combinations of realm (marine, terrestrial, freshwater), latitude (polar, temperate, tropical), and taxonomic-habitat group, allowing us to characterize variation in biodiversity change for taxon groups across broad latitudinal bands within each realm. The different geographic structure meant that this model included 271 studies when we applied our threshold of three cell-level time series per realm-latitude-taxon group (38). Results of all models were qualitatively consistent, both in terms of the overall trends they estimated and at the lowest levels (i.e., at the study and cell levels: fig. S3), suggesting that our inferences are largely robust to differences in how we searched for geographic patterns. Additionally, we found our results relatively insensitive to the heterogeneity in temporal extent of the data and did not detect systematic effects of the number of years sampled, temporal duration,

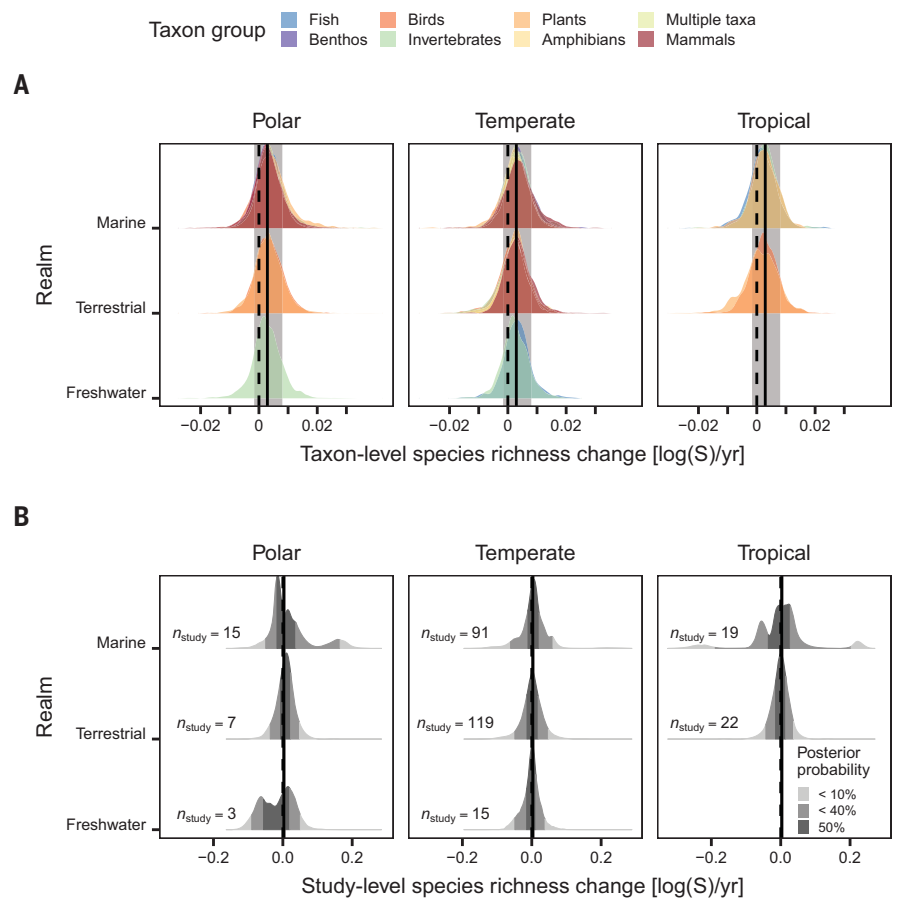


Fig. 2. Posterior distributions of species richness change for the realm-latitude-taxon model. The overall trend in assemblage richness change (solid vertical bar) does not differ from zero (gray shading depicts the 90% credible interval) for the realm-latitude-taxon study model. **(A)** Density ridges of the taxon-level slope coefficients (color represents the taxonomic group). **(B)** Density ridges of the posterior distributions of the study-level slope coefficients within a given combination of realm and latitudinal band estimated with the realm-latitude-taxon model.

start year, or the initial species richness on the estimates of rates of change (figs. S7 and S8).

Our biome-taxon model results show that variation in biodiversity change is greater in the marine versus the terrestrial and freshwater realms. The overall average of richness change was not statistically distinguishable from zero globally or for any individual biome (Fig. 1). The magnitude of positive departures from the overall trend was greater among marine biomes (range of median biome departure: -0.0003 to 0.001 , $n = 33$; Fig. 1A) compared with terrestrial and freshwater biomes (-0.0007 to 0.0001 , $n = 15$; Fig. 1B), but richness trends did not vary substantially among biomes ($\sigma_{Biome} = 0.004$) or for taxon groups within biomes ($\sigma_{Biome-taxon} = 0.003$). Instead, the main level of variation was at the study level ($\sigma_{Biome-taxon-study} = 0.04$), where specific studies exhibited species richness increases or decreases of up to 20% per year in the marine realm and up to 10% in the terrestrial realm (fig. S9). Twenty-three

marine, five terrestrial, and two freshwater studies showed significant species richness losses, whereas 31 marine and nine terrestrial studies showed significant gains. These results were consistent with the realm-latitude-taxon model that showed change centered on zero for all latitudinal bands (Fig. 2A), with the greatest variation observed in the marine realm at the study level, particularly in polar and tropical latitudes (Fig. 2B). Data limitations from tropical systems remain in our assemblage time-series data (e.g., no tropical freshwater assemblages), precluding some direct comparison between realms (see also fig. S1). The high rates of change that we observed in the marine tropics (Fig. 2B) are consistent with predictions that tropical marine species will be relatively sensitive to extreme heat events because they are closer to their physiological limits (20, 21), in addition to overexploitation, pollution, and other threats occurring in the marine tropics (36).

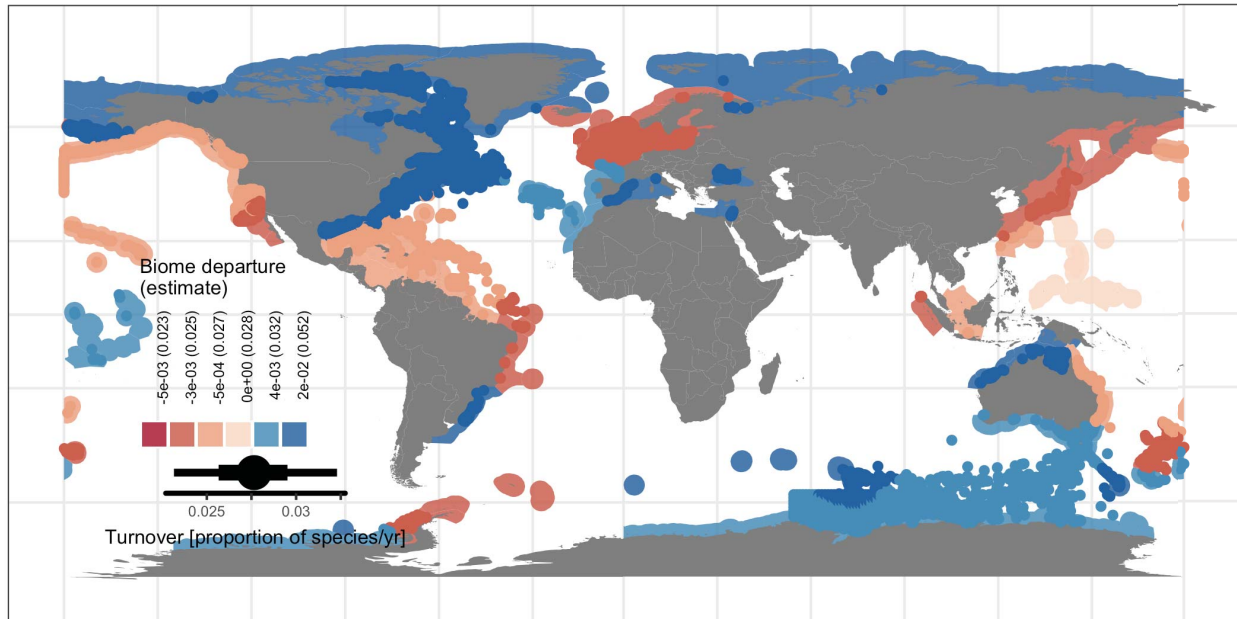
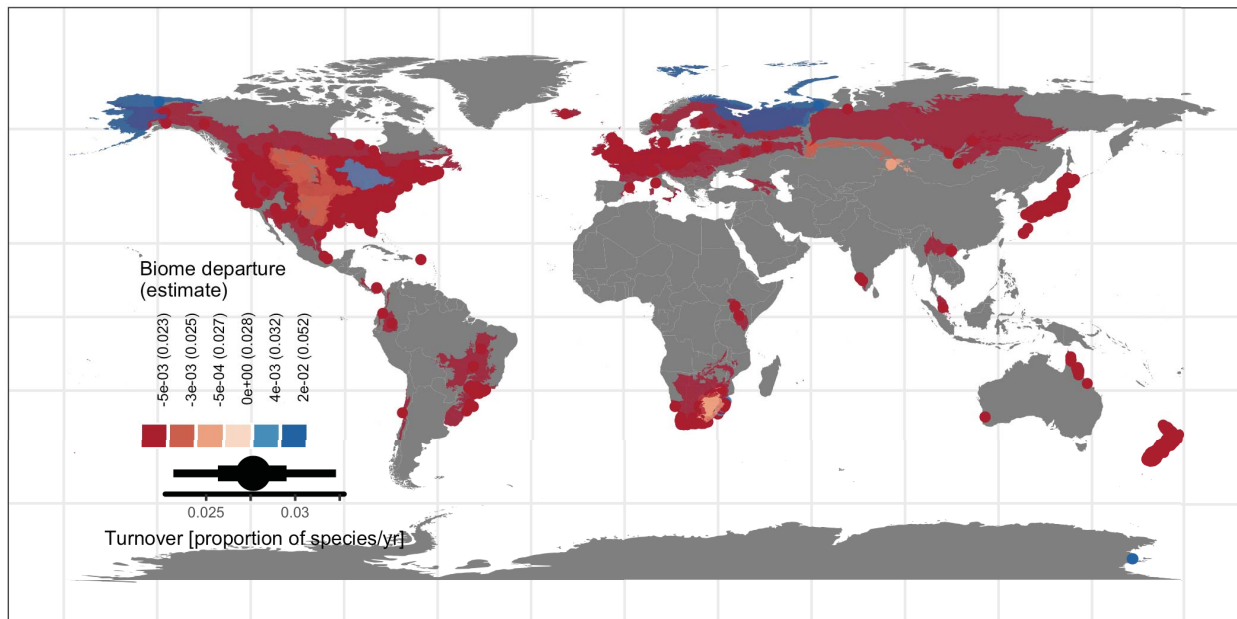
A Marine**B Terrestrial and freshwater**

Fig. 3. Species turnover component maps showing departures from the overall trend for marine and terrestrial biomes. Assemblages across the globe are experiencing high rates of species replacement (median ~28% of species replaced per decade). Shading represents positive (blue; faster turnover than average) and negative (red; slower turnover than average) departures from the overall trend for each biome; numbers in the inset denote the departure and the biome level (overall + departure) estimate in parentheses. **(A)** Rates of arriving species replacing original species have both faster (blue) and slower

rates of turnover (red) from the overall trend in marine biomes, but included the biomes with the highest turnover rates: the 90% credible intervals in the warm temperate northwestern Atlantic, warm temperate southwestern Atlantic, and the northwest Australian shelf biomes were greater than the overall trend, whereas **(B)** terrestrial and freshwater biomes have mostly slower rates of turnover than the overall trend (red shading), and the 90% credible intervals for temperate broadleaf and mixed forests, temperate conifer forests, and tropical and subtropical moist broadleaf forests were lower than the overall trend.

To examine changes in species composition, we partitioned total Jaccard dissimilarity, calculated as the dissimilarity between the initial year and each subsequent year of a time series, into the additive components of turnover and nestedness (45). These trends describe directional compositional change relative to the initial assemblage, and the decomposition determines whether changes in community composition were caused by the original species in assemblages being replaced by other species (turnover) or if assemblages were becoming smaller subsets of themselves or growing to include additional species alongside the original species (nestedness). Overall, we found that rates of turnover were positive and much greater (0.028; 90% credible interval: 0.023 to 0.032; Fig. 3) than the rates of change in nestedness (0.006; 0.006 to 0.007; fig. S12). Compositional change was dominated by species replacement within assemblages, with ~28% of species being replaced per decade. Variation at the biome level was much greater

for turnover ($\sigma_{\text{Biome}} = 0.01$) compared with species richness, resulting in stronger geographic patterns and revealing further differences between marine and terrestrial realms (Fig. 4). Three marine biomes (warm temperate northwest and southwest Atlantic and the northwest Australian shelf) had rates of turnover greater than the overall trend. By contrast, three terrestrial biomes (temperate broadleaf and mixed forests, temperate conifer forests, and tropical and subtropical moist broadleaf forests) had rates of turnover slower than the global trend, and most terrestrial biomes showed negative departures from the global average (Fig. 3B). Positive departures from the overall trend in terrestrial and freshwater biomes were found in aquatic systems: large lakes, mangroves, and polar freshwaters. These trends of directional compositional change are highly unlikely to have arisen simply from random assemblages being drawn from relatively constant regional species pools. Simulations show that such a process has a

median slope of zero for both turnover and nestedness change (38) (fig. S10). Additionally, we found that higher rates of compositional change in marine and freshwater biomes were associated with a higher proportion of assemblages undergoing complete turnover (fig. S13) and were robust to our choice of error distribution (fig. S14) and whether comparisons were made with the initial assemblage or between assemblages at consecutive time points (38) (figs. S15 and S16).

Linking richness and composition change

To examine the relationship between changes in species richness and changes in composition, we plotted the dominant component of composition change (turnover or nestedness) for each biome-taxon-study combination against species richness change (46) (Fig. 5, A and B). When turnover is the dominant component, this relationship shows how quickly different species are replacing original species and whether these arrivals are associated with changes to the number of species. At the study level, rates of turnover exceeded nestedness change for >97% of biome-taxon-study combinations (313/321; Fig. 5, B and C). Among these studies, ~23% (57/313) exhibited trends different from zero for both turnover and species richness rates (Fig. 5C), with a relatively balanced distribution of 23 cases of species richness losses and 34 cases with gains. We did not adjust for multiple comparisons, though they are less of a problem when comparing partially pooled estimates from hierarchical models (47). When nestedness is the dominant component, this relationship shows how fast assemblages are changing to become smaller subsets or growing to include additional species alongside the species initially observed. Among the eight biome-taxon-study combinations where nestedness exceeded turnover change (8/321; Fig. 5, B and C), only two showed rates of nestedness and richness trends different from zero, with one losing and one gaining species. Our combined results for turnover and species richness change support recent studies reporting that different components of biodiversity change, such as composition shifts and species richness, are largely uncoupled (5, 7–9). In fact, we found that high rates of turnover were associated with the full spectrum of richness changes.

Discussion

Compositional change dominated by species turnover is the most conspicuous and prevalent form of biodiversity change across the globe and was characterized by strong geographic structure. Only marine biomes were found to have faster rates of compositional change than the overall trend, whereas only terrestrial biomes were observed to trail the overall compositional trend. Moreover, marine studies exhibited greater

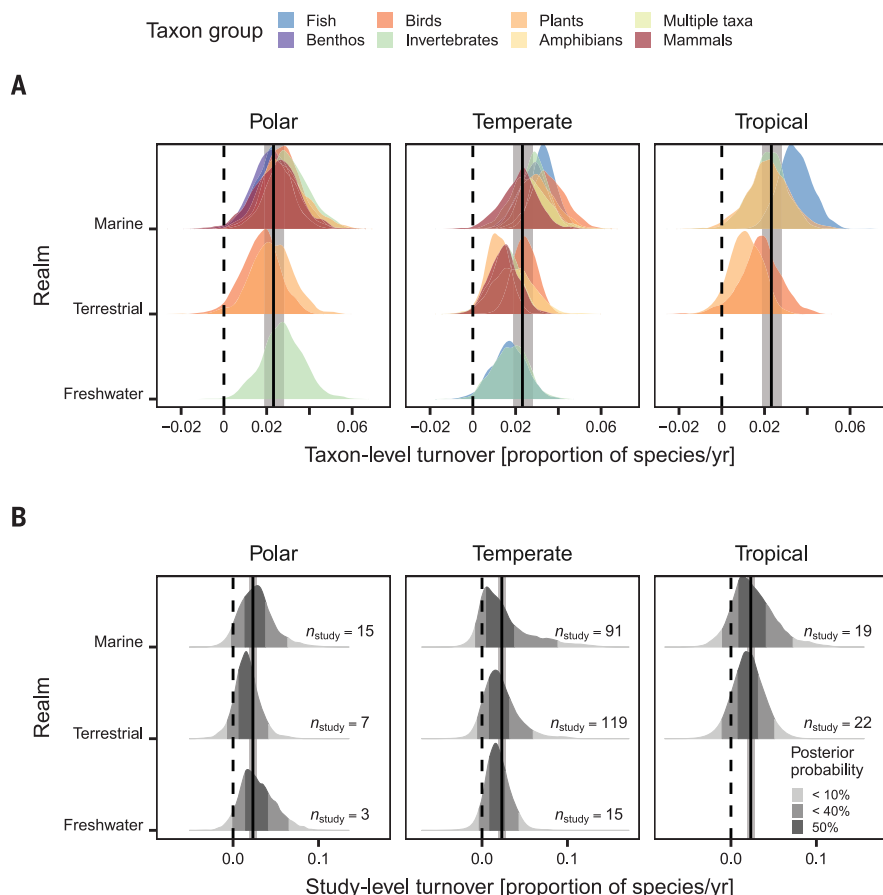


Fig. 4. Posterior distributions of turnover for the realm-latitude-taxon model. The overall trend in turnover change per year is greater than zero (solid black line; gray shading depicts the 90% credible interval) for the realm-latitude-taxon-study model. **(A)** Density ridges of the taxon-level slope coefficients (color represents the taxonomic group). **(B)** Density ridges of the study-level slope coefficients within a given combination of realm and latitudinal band estimated with the realm-latitude-taxon model.

variation in rates of compositional turnover. We also found that most studies, across all biomes and realms, showed considerable replacement of species through time without associated species richness changes. This finding, using the largest global dataset to date, is consistent with the assertion that species richness trends are often uncoupled from species replacement, and thus are insufficient alone for fully capturing how biodiversity might change (9). The consistent

pattern of species replacement is likely underpinned by a diverse suite of drivers affecting different study sites, regions, and realms. Such reorganization independent of changes to the number of species is also consistent with the presence of regulatory mechanisms for species richness. Community regulation of species richness is widespread (48) and may be driven, for example, by shared resources (24). Contemporary pressures such as introduced species (29–31), replacement of localized spe-

cialists by widespread generalists (49, 50), range shifts in response to environmental change (22, 51), or local warming (52, 53) may also help to explain our finding of widespread composition change associated with variable richness change at the study level.

Rates of species richness change and turnover were higher in absolute magnitude and more variable in the marine realm, with maximum turnover rates in marine biomes twice those observed for terrestrial biomes. Higher rates of turnover in the marine realm are consistent with predictions for species responses on the basis of greater sensitivities to increased temperatures (20, 21) coupled with fewer barriers to dispersal in marine systems (22, 23), though attribution to specific drivers is beyond the scope of analyses presented here. Further, if assemblages are more spatially heterogeneous in marine compared with terrestrial systems, then this too may contribute to our finding of higher temporal turnover in marine assemblages (54). Although we did not find strong contrasting trends for specific taxonomic groups within biomes, environmental differences between the realms (unrelated to dispersal and connectivity) or life history differences among taxa could also underpin some of the patterns in turnover detected at the realm and regional scales. For example, in the temperate marine realm, mammals had lower turnover than invertebrates, which is consistent with other findings showing that long-lived taxa exhibit less rapid temporal turnover (54).

Amid widespread variation in biodiversity trends, we found that tropical marine regions have a higher proportion of studies exhibiting biodiversity change at the extremes of richness gains, losses, and turnover (Figs. 2B and 4 and fig. S9). Hence, although we find higher magnitude changes in the tropics, this result contrasts with our prediction that we would find mostly richness losses. The tropics, which harbor most of the biological diversity on the planet, are also generally considered to be the place where biodiversity is the most threatened (36). Moreover, in the context of climate change, there are likely fewer species available to replace those species lost in tropical zones that have entered no-analog warm-temperature conditions (55, 56). If these trends are maintained, then this could lead to a market restructuring of biodiversity, with potentially severe consequences for ecosystem functioning across biomes and changes to the latitudinal diversity gradient, substantially altering the planet's biogeography. However, such a pattern of biotic attrition would be expected to be accompanied by a larger contribution of the nestedness component to community compositional change than we found here (Fig. 5 and fig. S12). BioTIME includes relatively few tropical datasets (37),

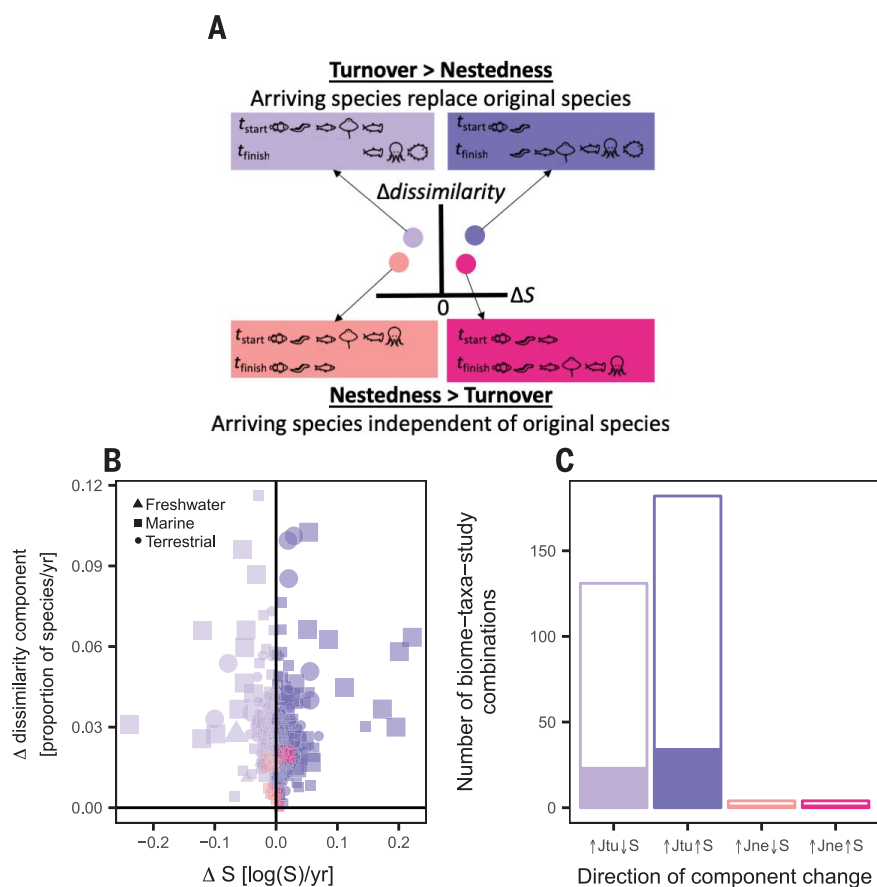


Fig. 5. Conceptual and empirical relationships between changes in species richness and changes in species composition. (A) Conceptual model relating the turnover and nestedness components of species composition change ($\Delta dissimilarity$) to changes in species richness (ΔS). When the turnover component is larger than the nestedness component, different species entering assemblages replace the original species (purple shaded boxes). Conversely, when the nestedness component is larger than the turnover component, some original species of the assemblage remain and the numbers of different species entering the assemblage are largely independent of the original species (pink shaded boxes). The change in species richness documents the net change in the numbers of species in the assemblage (and ignores their identity as either original or additional species). (B) Scatterplot showing the dissimilarity trend as a function of the species richness trend. Each point represents a biome-taxon-study level estimate, point shape indicates realm, and larger points indicate that both composition and richness trends differed from zero with 90% probability (credible intervals are not shown for clarity). (C) Number of studies for each combination of change in species richness and species composition measured as either the turnover (Jtu) or nestedness (Jne) component of Jaccard dissimilarity. Filled sections of each bar represent the number of biome-taxon-study combinations where the 90% credible intervals for both species richness (S) and composition change (Jtu or Jne) did not overlap zero.

despite being the largest compilation of biodiversity time series currently available, and further data collection in these areas is needed to be able to confidently assess trends in the tropics. Furthermore, biodiversity monitoring overall is lacking for many regions of the planet, e.g., the deep ocean. The geographical variation that we uncovered highlights the critical importance of improving the spatial coverage of biodiversity monitoring to better estimate global biodiversity change.

We identified hotspots of biodiversity change, that is, key areas that represent extremes for biodiversity trends. As conservation has moved toward systematically identifying regions in need of protection (57), global maps of conservation priorities according to different criteria have been developed (16, 58). Our study provides an important criterion for targeting conservation action: a global map of current rates of biodiversity change. In addition to the marine tropics, marine biomes in the western Atlantic and northwest Australia are undergoing rates of replacement higher than the global average. Therefore, these regions are currently undergoing the most substantial rates of change and should be prioritized for reactive conservation measures. By contrast, several forest biomes (e.g., temperate broadleaf and mixed, temperate conifer) have slower rates of replacement than the global average. Because these regions appear to be undergoing less change, we speculate that proactive conservation measures are likely to be more appropriate. Specific conservation actions always need to be tailored to the locations and taxa, but our study provides the global and regional context in which individual locations are immersed.

The global heterogeneity in biodiversity change is underpinned by geographic variation. We find that spatial variation in species gains and losses is greater than taxonomic variation. This spatial variation suggests that statements about biodiversity loss need to be conditional on context and location. On average, local species richness change across the globe does not differ from zero, but there are many locations gaining or losing species. Species replacement is ubiquitous and also spatially structured. Determining whether the spatial variation uncovered is related to differences among communities in their degree of exposure and vulnerability to drivers of biodiversity change is an important next step. Nevertheless, our results show that, although the entire planet is undergoing biodiversity change, the direction and magnitude of change differs across geographic regions.

REFERENCES AND NOTES

1. S. L. Pimm *et al.*, *Science* **344**, 1246752 (2014).
2. V. Proença, H. M. Pereira, "Comparing extinction rates: Past, present and future," in *Encyclopedia of Biodiversity*, S. A. Levin, Ed. (Elsevier, ed. 2, 2013), pp. 167–176.
3. H. M. Pereira *et al.*, *Science* **330**, 1496–1501 (2010).
4. J. García Molinos *et al.*, *Nat. Clim. Chang.* **6**, 83–88 (2016).
5. M. Dornelas *et al.*, *Science* **344**, 296–299 (2014).
6. M. Vellend *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19456–19459 (2013).
7. S. R. Supp, S. K. M. Ernest, *Ecology* **95**, 1717–1723 (2014).
8. R. Elahi *et al.*, *Curr. Biol.* **25**, 1938–1943 (2015).
9. H. Hillebrand *et al.*, *J. Appl. Ecol.* **55**, 169–184 (2018).
10. E. R. Pianka, *Am. Nat.* **100**, 33–46 (1966).
11. M. L. Rosenzweig, *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
12. A. Gonzalez *et al.*, *Ecology* **97**, 1949–1960 (2016).
13. M. Vellend *et al.*, *Ecology* **98**, 583–590 (2017).
14. L. Tydecks, J. M. Jeschke, M. Wolf, G. Singer, K. Tockner, *PLOS ONE* **13**, e0199327 (2018).
15. C. Meyer, P. Weigelt, H. Kreft, *Ecol. Lett.* **19**, 992–1006 (2016).
16. T. M. Brooks *et al.*, *Science* **313**, 58–61 (2006).
17. C. J. Ellis, *Syst. Biodivers.* **9**, 307–317 (2011).
18. B. S. Halpern *et al.*, *Nat. Commun.* **6**, 7615 (2015).
19. D. E. Bowler *et al.*, *bioRxiv* 432880 [Preprint]. 2 October 2018.
20. J. M. Sunday, A. E. Bates, N. K. Dulvy, *Nat. Clim. Chang.* **2**, 686–690 (2012).
21. M. L. Pinsky, A. M. Eikeset, D. J. McCauley, J. L. Payne, J. M. Sunday, *Nature* **569**, 108–111 (2019).
22. E. S. Poloczanska *et al.*, *Nat. Clim. Chang.* **3**, 919–925 (2013).
23. R. M. May, *Philos. Trans. R. Soc. London B Biol. Sci.* **343**, 105–111 (1994).
24. J. H. Brown, S. K. M. Ernest, J. M. Parody, J. P. Haskell, *Oecologia* **126**, 321–332 (2001).
25. M. Dornelas *et al.*, *Ecol. Lett.* **22**, 847–854 (2019).
26. T. Newbold *et al.*, *Nature* **520**, 45–50 (2015).
27. R. K. Colwell, G. Brehm, C. L. Cardelús, A. C. Gilman, J. T. Longino, *Science* **322**, 258–261 (2008).
28. M. van Kleunen *et al.*, *Nature* **525**, 100–103 (2015).
29. D. F. Sax, S. D. Gaines, *Trends Ecol. Evol.* **18**, 561–566 (2003).
30. M. Winter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21721–21725 (2009).
31. M. Winter *et al.*, *Glob. Ecol. Biogeogr.* **19**, 332–342 (2010).
32. C. Parmesan, G. Yohe, *Nature* **421**, 37–42 (2003).
33. J. Lenoir, J. C. Svenning, *Ecography* **38**, 15–28 (2015).
34. L. Fahrig, *Annu. Rev. Ecol. Evol. Syst.* **48**, 1–23 (2017).
35. J. W. Williams, S. T. Jackson, J. E. Kutzbach, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5738–5742 (2007).
36. J. Barlow *et al.*, *Nature* **559**, 517–526 (2018).
37. M. Dornelas *et al.*, *Glob. Ecol. Biogeogr.* **27**, 760–786 (2018).
38. See supplementary materials.
39. R. Barnes, *dgggridR*: Discrete global grids. R package version 2.0.3 (2018); <https://CRAN.R-project.org/package=dgggridR>.
40. D. M. Olson, E. Dinerstein, *Ann. Mo. Bot. Gard.* **89**, 199–224 (2002).
41. M. D. Spalding *et al.*, *Bioscience* **57**, 573–583 (2007).
42. R. Abell *et al.*, *A.I.B.S. Bull.* **58**, 403–414 (2008).
43. The Nature Conservancy, *Terrestrial Ecoregions Dataset: Vector Digital Data*. (2009); http://maps.tnc.org/gis_data.html.
44. D. M. Olson *et al.*, *Bioscience* **51**, 933–938 (2001).
45. A. Baselga, *Glob. Ecol. Biogeogr.* **19**, 134–143 (2010).
46. C. Albouy, F. Guilhaumon, M. B. Araújo, D. Moullot, F. Leprieux, *Glob. Chang. Biol.* **18**, 2995–3003 (2012).
47. A. Gelman, J. Hill, M. Yajima, *J. Res. Educ. Eff.* **5**, 189–211 (2012).
48. N. J. Gotelli *et al.*, *Sci. Adv.* **3**, e1700315 (2017).
49. I. Le Viol *et al.*, *Biol. Lett.* **8**, 780–782 (2012).
50. L. Alvarez-Filip, M. J. Paddack, B. Collen, D. R. Robertson, I. M. Côté, *PLOS ONE* **10**, e0126004 (2015).
51. R. D. Batt, J. W. Morley, R. L. Selden, M. W. Tingley, M. L. Pinsky, *Ecol. Lett.* **20**, 1148–1157 (2017).
52. P. H. Flanagan, O. P. Jensen, J. W. Morley, M. L. Pinsky, *Ecography* **42**, 214–224 (2019).
53. H. Hillebrand, J. Soininen, P. Snoeij, *Glob. Change Biol.* **16**, 1181–1193 (2010).
54. S. L. Collins *et al.*, *Ecology* **99**, 858–865 (2018).
55. M. T. Burrows *et al.*, *Nature* **507**, 492–495 (2014).
56. M. L. Pinsky, B. Worm, M. J. Fogarty, J. L. Sarmiento, S. A. Levin, *Science* **341**, 1239–1242 (2013).
57. C. R. Margules, R. L. Pressey, *Nature* **405**, 243–253 (2000).
58. N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, J. Kent, *Nature* **403**, 853–858 (2000).
59. S. A. Blowes *et al.*, *Code for: The geography of biodiversity change in marine and terrestrial assemblages*. Zenodo (2019); <https://doi.org/10.5281/zenodo.1473860>.

ACKNOWLEDGMENTS

We thank all of the authors who contributed their data or gave us permission to include it in this study. M. O'Connor helped in securing the funding for the working group and contributed to discussions and editing the manuscript drafts. Many working group members contributed to projects in various roles across multiple meetings and were influential in shaping the discussions and ideas that led to this paper. Those not listed as authors here include: M. O'Connor, J. Sarmiento Cabral, J. Dunic, and R. Elahi. We also thank C. Krause and the UFZ administrative and support staff of the High-Performance Computing Cluster EVE, a joint effort of the Helmholtz Centre for Environmental Research (UFZ) and iDiv, for access to EVE. S.A.B. thanks J. Belmaker and the Marine Ecology & Biodiversity lab at Tel Aviv University. S.R.S. thanks M.A.S. and H.A.S. for bookending the ideas and research that gestated to become this manuscript. **Funding:** This work was supported by funding to the sChange working group through sDiv, the synthesis center of iDiv, the German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118). S.A.B., H.B., J.M.C., J.H., and M.W. were supported by the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig. S.R.S. was supported by U.S. National Science Foundation grant 1400911. LHA was supported by Fundação para a Ciência e Tecnologia, Portugal (POPH/FSE SFRH/BD/90469/2012), and by the Jane and Aatos Erkko Foundation. M.D. was supported by a Leverhulme Trust Fellowship. A.E.M., F.M., and M.D. were supported by ERC AdG BioTIME 250189 and PoC BioCHANGE 727440. A.G. is supported by the Liber Ero Chair in Biodiversity Conservation. **Author contributions:** M.D. conceived the project. S.A.B., S.R.S., and M.D. led the development of the project, assisted with data analysis and interpretation, and wrote the first draft of the manuscript. S.A.B. and S.R.S. collaborated on the core data preparation and coding in R. S.A.B. designed the analytical models and prepared the figures for the manuscript. F.M. managed the BioTIME database, queried it for the analysis, and provided help with figures. All authors contributed to the sDiv working group that conceived this project, to key discussions that led to the design of the study, and to revising the paper and preparing and approving it for publication. M.D. and S.R.S. secured the funding from sDiv and co-led the sChange working group (with M. O'Connor) that initiated the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The time series analyzed were from 332 references found in the BioTIME dataset and in other studies that were used with permission. Approximately 92% (306 references) of the biodiversity studies analyzed here are available as part of the published BioTIME Database (28). The data are openly available and can be accessed on Zenodo (<https://doi.org/10.5281/zenodo.2602708>) or through the BioTIME website (<http://biotime.st-andrews.ac.uk/>). The remaining 8% (26 references) of biodiversity studies analyzed were used with permission. Data from these studies are published and publicly available outside of the BioTIME database and are noted with an asterisk in the supplementary materials (table S1), with the original reference and online source for download. Code is available in an online archive at Zenodo (59).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/339/suppl/DC1
Materials and Methods
Figs. S1 to S18
Table S1
References (60–481)

23 November 2018; accepted 10 September 2019
10.1126/science.aaw1620

EARTHQUAKES

Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence

Zachary E. Ross^{1*}, Benjamín Idini¹, Zhe Jia¹, Oliver L. Stephenson¹, Minyan Zhong¹, Xin Wang¹, Zhongwen Zhan¹, Mark Simons¹, Eric J. Fielding², Sang-Ho Yun², Egill Hauksson¹, Angelyn W. Moore², Zhen Liu², Jungkyo Jung²

A nearly 20-year hiatus in major seismic activity in southern California ended on 4 July 2019 with a sequence of intersecting earthquakes near the city of Ridgecrest, California. This sequence included a foreshock with a moment magnitude (M_w) of 6.4 followed by a M_w 7.1 mainshock nearly 34 hours later. Geodetic, seismic, and seismicity data provided an integrative view of this sequence, which ruptured an unmapped multiscale network of interlaced orthogonal faults. This complex fault geometry persists over the entire seismogenic depth range. The rupture of the mainshock terminated only a few kilometers from the major regional Garlock fault, triggering shallow creep and a substantial earthquake swarm. The repeated occurrence of multifault ruptures, as revealed by modern instrumentation and analysis techniques, poses a formidable challenge in quantifying regional seismic hazards.

Over the past three decades, an increasing number of well-documented earthquakes have ruptured multiple faults (1–5), highlighting the importance of incorporating this geometric complexity into models of seismic hazard. On 4 July 2019, a sequence of damaging earthquakes began near the city of Ridgecrest, California, and activated a complex fault network, further illustrating the need to understand how multiple faults can rupture in a single earthquake.

The Ridgecrest sequence included a foreshock with a moment magnitude (M_w) of 6.4 followed by an M_w 7.1 mainshock nearly 34 hours later (Fig. 1). These dominantly strike-slip earthquakes occurred on largely unmapped faults that cumulatively extend more than 75 km in length. Shaking from these events was felt throughout portions of California, Nevada, and even Arizona. Damage was primarily concentrated in the towns of Ridgecrest and Trona. The severity of damage was limited, as the area is sparsely populated. However, substantial portions of the ruptures occurred within the boundaries of the Naval Air Weapons Station at China Lake and caused major damage to facilities there.

Seismic activity occurred within the Little Lake fault zone (LLFZ) and nearby Airport Lake fault zone, both of which have a long history of activity (Fig. 1), including multiple sequences with M_w of >5 between 1982 and 1996 (6). The LLFZ is accumulating right-lateral strain of about 1 mm/year (7) and is bounded to the southeast by the Garlock fault, a 260-km-long left-lateral strike-slip fault capable of producing M_w ~7.8 earthquakes (8). The Coso geothermal area abuts the LLFZ to the northwest and is the site of one of the largest geothermal

power plants in the United States. Numerous earthquake swarms have occurred in this area (9). To the northwest of Coso is the southern terminus of the 1872 M_w 7.5 Owens Valley earthquake (10) (Fig. 1), which is among the largest historical earthquakes in California and was responsible for 27 deaths.

In the time since the last major earthquake in California, the quality of seismic and geodetic data available has improved substantially. The Southern California Seismic Network has nearly doubled in size, Global Positioning System (GPS) stations have been installed throughout the region, and interferometric synthetic aperture radar (InSAR) data are available at unprecedented levels of quality with shorter times between acquisitions. Here, we analyze these datasets with state-of-the-art techniques, including a high-resolution seismicity catalog derived from template matching, detailed multifault inversions of geodetic data with nonplanar geometry, and subevent modeling of the largest events in the sequence. The observations indicate that the Ridgecrest sequence ruptured an astonishingly complex network of intertwined orthogonal faults operating over a broad range of length scales. They also show that the Garlock fault experienced shallow triggered postseismic creep and a large earthquake swarm; the role this major tectonic structure plays in limiting the southern extent of faulting and seismicity is yet to be understood.

The 2019 Ridgecrest sequence

The Ridgecrest mainshock was preceded by a prominent foreshock sequence that began with numerous events of local magnitude (M_L) ~0 (Fig. 1). The seismicity rates escalated rapidly, with an M_w 4 event 2 hours later and an M_w 6.4 earthquake shortly thereafter. Over the first 21 days of the sequence, more than 111,000 earthquakes with M of >0.5 occurred, including 70 events with an M of >4.

To resolve the geometry of the fault zone that ruptured during the Ridgecrest sequence, we produced a comprehensive relocated seismicity catalog with template matching that had nominal relative resolution of ~100 m horizontally and 350 m vertically (11) (Fig. 2). We paired this catalog with a damage proxy map (DPM) (11) (Fig. 3), extracted from satellite InSAR coherence data, that identifies regions where radar-scattering properties of the ground have changed substantially relative to their inherent background rates of change (12). The fault zone exhibits pervasive orthogonal faulting over multiple length scales with notable geometric complexity (Figs. 2 and 3). The largest scale is for a roughly 55-km-long northwest-striking structure with an undulating geometry that is orthogonally cross-cut by a ~15-km-long fault. The larger of these was the primary structure that ruptured during the M_w 7.1 earthquake, while the smaller structure was the largest that ruptured during the M_w 6.4 event (Fig. 3).

Around the northwest rupture terminus, the main fault dips steeply to the southwest before transitioning to a northeast dip of roughly 70° on the segment of the fault southeast of the M_w 7.1 hypocenter. Toward the southeast part of the rupture, the largest fault bifurcates into two subparallel strands that are separated by 7 km and that continue for about 12 km each (Figs. 2 and 3). Near the main intersection of the M_w 7.1 and M_w 6.4 ruptures, there are at least 20 orthogonal faults that appear as lineations in the seismicity and features in the DPM. The lineations often localize to a width of roughly 100 to 200 m, which is comparable to the uncertainty of the hypocenter locations (11). The general absence of orthogonal faulting in the northwest section of the rupture may indicate that this part of the fault zone is slightly further along in the evolutionary process. Numerous horsetail structures are oriented at an oblique angle to the main rupture trace at the northwest rupture terminus.

Nearly all of the geometric features described are present in the seismicity as well as the InSAR surface deformation, indicating that they are representative of the fault structure over the entire seismogenic zone. As defined by seismicity, the bottom of the seismogenic zone lies around 10 km at its deepest extent, with the maximum depth slowly decreasing to about 5 km toward the southeast (Fig. 2). Such a shallow seismogenic zone is also present in the Salton Trough region of California and is often attributed to the elevated heat flow (13); similar heat flow anomalies exist near the Ridgecrest sequence (14) and may control the boundaries of the brittle-ductile transition zone.

On the basis of kinematic subevent inversion of seismograms from the dense regional seismic network and global seismic stations, the M_w 6.4 earthquake had a duration of about

¹Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA. ²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA.
*Corresponding author. Email: zross@caltech.edu

12 s, with three subevents best explaining the data (Fig. 4). These three subevents coincide with at least three faults. The 6-km-long northwest-trending fault slipped first, with an equivalent M_w 6.1 (subevent E1). This is consistent with the hypocenter of this event, which is located about 2 km northwest of the long southwest-trending fault. Thus, a rupture propagated over a short southwest-trending fault with only about 5 km of surface break, yet the equivalent magnitude for this segment, was the M_w 6.2 subevent (E2). Our analysis of this subevent indicates that it is required by the data and that the magnitude is well resolved (figs. S1 to S4). However, this subevent may also include slip on an adjacent northwest-trending orthogonal fault. Because the seismicity on the E2 segment extends across the main fault of the main-shock (Fig. 2), part of the slip probably occurred on the other side, but only at depth. The rupture then jumped to a larger southwest-trending fault that is about 15 km long, producing again an equivalent seismic moment to an M_w 6.2 earthquake (E3) and a large surface rupture.

Additional insights into the rupture process of the M_w 6.4 earthquake are given by a joint inversion of the geodetic data spanning both events and constraints provided by the relocated seismicity (Fig. 5 and fig. S10). On the faults closest to the three kinematic subevents, there is appreciable slip in our model. This event nucleated at the intersection between two faults trending northwest and southwest (Fig. 5). The 15-km-long southwest-oriented left-lateral fault that aligns with subevent E3 (Fig. 4) produced a maximum of about 5 m slip. The large slip patches on these faults are coincident with regions of decreased aftershock density (fig. S10).

Our kinematic subevent model of the M_w 7.1 mainshock shows that the rupture lasted for about 22 s and can be explained by four subevents (Fig. 4). The initial 5 s of the rupture had little moment, whereas the subevent with the largest moment (subevent E1, M_w 6.9) occurred close to the hypocenter between 5 and 10 s. This subevent appears to have ruptured bilaterally, given its large seismic moment. Centroid locations of the later subevents propagated very slowly to the southeast, suggesting a unilateral rupture. The final subevent, E4 (at 20 s), occurred ~25 km southeast of the first subevent E1 (at 7 s), indicating a slow average rupture velocity of roughly 2 km/s.

The M_w 7.1 earthquake nucleated about 10 km to the northwest of the M_w 6.4 event and ruptured a fault network with a cumulative length of about 65 km. Most of the slip occurred near the hypocenter (Fig. 5), where an ~30-km-long patch produced a maximum of about 9 m slip at the scales resolved by our model. This part of the rupture is notably deficient in aftershocks relative to the rest of the rupture (fig. S10), a feature that is common to large slip

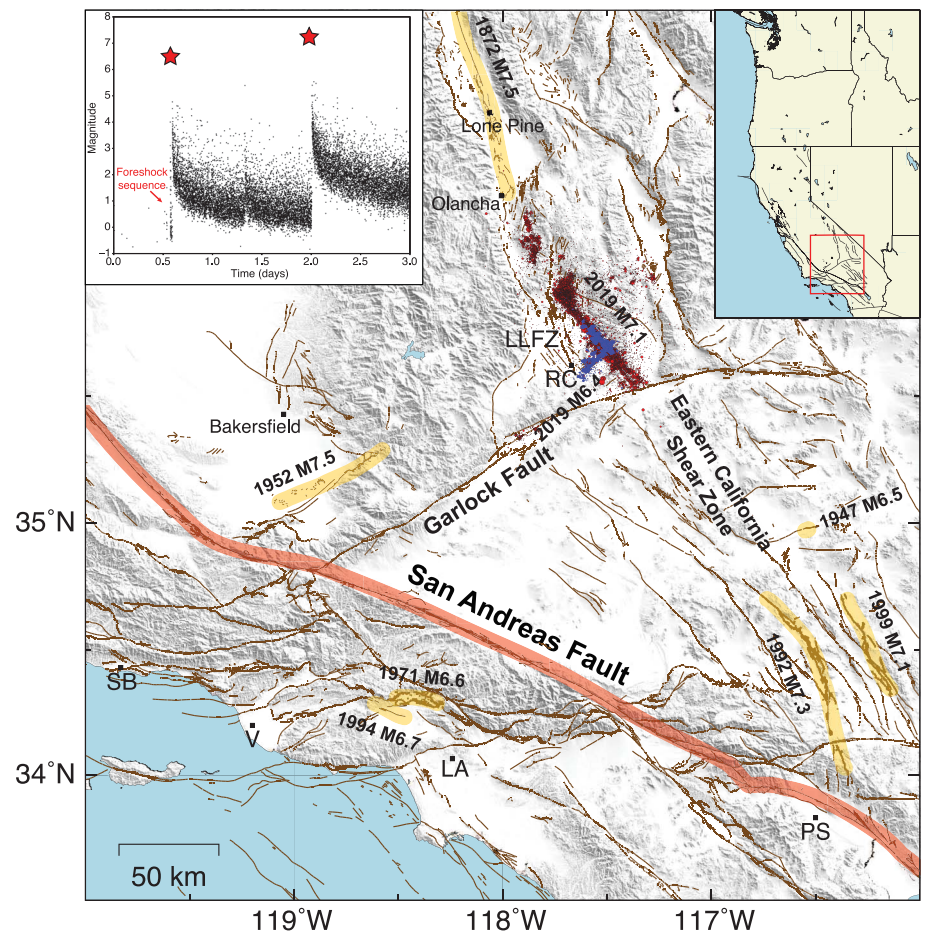


Fig. 1. Overview of study area. Mapped faults are indicated in brown. Ridgecrest sequence epicenters are shown as blue dots for time period up to the M_w 7.1 mainshock and as red dots following the mainshock. Santa Barbara, Los Angeles, Ventura, Palm Springs, and Ridgecrest and the LLFZ are denoted SB, LA, V, PS, RC, and LLFZ, respectively. Inset shows magnitude-time evolution of the first three days of the sequence. Note the catalog incompleteness during the first few days after the largest events. The M_w 6.4 event was preceded by a prominent foreshock sequence.

regions of many other earthquakes (15). Slip in the model has a maximum depth that is close to that of the seismicity (~10 km). We found that both strands of the bifurcation ruptured for about 12 km, terminating less than 5 km from the Garlock fault. These observations agree with the extent of seismicity. There is evidence from the seismicity for one or more additional subparallel faults on the southwest side of these two (Fig. 2).

The faults that ruptured during the M_w 7.1 event appear to be distinct from those that produced the M_w 6.4 event, which likely indicates that no fault ruptured the same area twice. The initial portion of the M_w 6.4 rupture propagated to the northwest about 6 km and terminated about 4 km from the eventual M_w 7.1 hypocenter (Fig. 4 and fig. S7). This 4-km gap was progressively filled by a series of moderate-sized earthquakes in the 34 hours after the M_w 6.4 event, which suggests that this portion of the fault acted as a barrier through which the

M_w 6.4 rupture was unable to propagate (Fig. 4 and fig. S7). However, from the DPM (Fig. 3B), it is clear that this 4-km-long segment did rupture the surface at some point during the sequence, which most likely occurred during the M_w 7.1 event. The M_w 7.1 event appears to have been triggered because steady seismic activity eroded away this barrier. Furthermore, from our modeling of the source process of the M_w 6.4 event, this foreshock either jumped across the primary fault that failed during the M_w 7.1 event or ruptured through it in an orthogonal direction, yet it was somehow unable to trigger this large fault at that time. Instead, the M_w 7.1 event nucleated more than 10 km away from this intersection after being weakened by a series of $M_w > 4$ earthquakes over 34 hours. The dynamic rupture conditions that could allow for such a scenario are not clear.

In addition to direct aftershocks that occurred in the LLFZ, the Ridgecrest sequence initiated three sizable earthquake swarms. One swarm

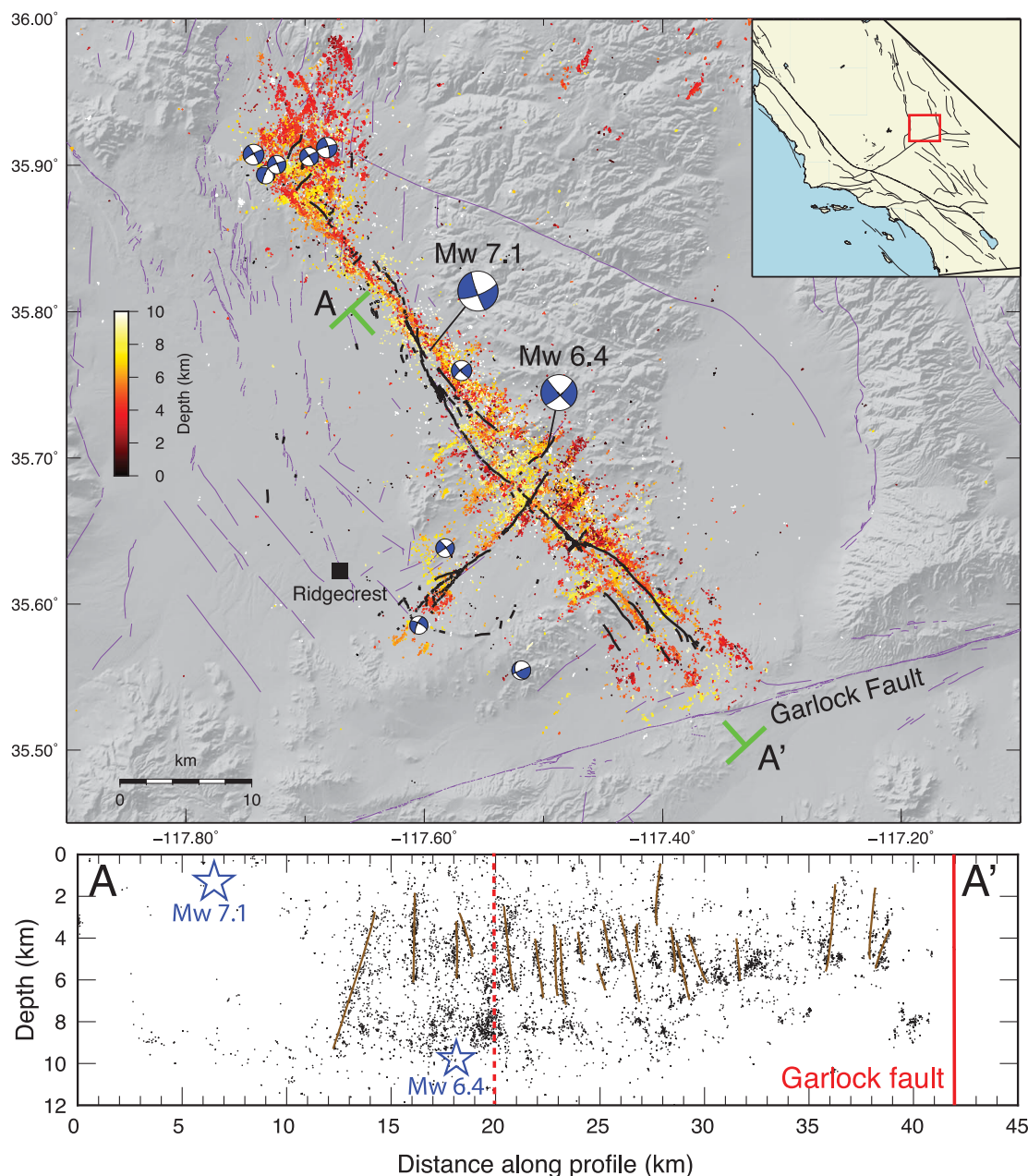


Fig. 2. Map view of Ridgecrest seismicity. Black lines indicate the surface trace of the fault (31), and purple lines indicate quaternary faults. Events with M_w of >4.5 are indicated by focal mechanisms (32). The fault network exhibits multiscale orthogonal faulting throughout the region, with a bifurcation to the southeast and horsetail faulting at the northwest terminus. The lower panel shows a seismicity cross section for events within 1 km of A-A' with interpreted faults drawn shown as brown lines. At least 20 orthogonal faults cut through this profile. The dashed red line indicates the surface trace of southwest-trending fault that ruptured in the M_w 6.4 event.

was activated on the Garlock fault by the M_w 6.4 event and locates in a pull-apart structure about 50 km southwest of the Ridgecrest sequence (fig. S9). More than 4000 events with an M of >0 occurred in this swarm during the first 3 weeks, the largest of which was M_L 3.2 (fig. S9). This area has seen only minor seismic activity over the past several decades. Another swarm began north of the Coso geothermal field (Fig. 1), where similar behavior was observed following the Landers earthquake (16), with a 15-km gap over the geothermal production area (17). A third swarm was triggered in Panamint Valley.

On the section of the Garlock fault south of the rupture terminus of the Ridgecrest main-

shock, InSAR data reveal an ~ 30 -km zone of left-lateral, triggered, shallow creep (Fig. 6). The largest surface offset is around 20 mm of relative motion in the satellite line-of-sight direction, directly on the bearing of the M_w 7.1 rupture. The narrow extent of the deformation field suggests that the creep is confined to the upper few hundred meters. This creep is similar to shallow creep induced by other major earthquakes (18). Although the Garlock fault has been seismically quiescent during the historic period, it has hosted numerous large earthquakes during the previous several thousand years (19). However, previous geodetic measurements have shown no measurable creep on the Garlock fault (20).

Discussion

Most of our knowledge about the structural architecture of fault zones comes from observations made on the surface, which consist primarily of geological mapping of faults and geodetic observations of coseismic deformation. At depth, our understanding of properties such as the geometry of fault zones is far less complete, with most evidence suggesting a general tendency for structural complications like damage zones to localize and simplify with depth (21). For the Ridgecrest earthquakes, we find that nearly every aspect of the surface geometry persists at depth, including the bifurcation of the M_w 7.1 rupture to the southeast, the horsetail faulting

at the northwest terminus, the multiple segments that ruptured during the M_w 6.4 rupture, and the rampant orthogonal faulting. Below a 6-km depth, the main faults that ruptured during the M_w 7.1 event may become slightly more planar. The smaller orthogonal faults produced events primarily with M of <3 , whereas events with M of >3 are mostly associated with the largest structures (fig. S15). These results show that precisely relocated seismicity catalogs are of sufficient resolution to constrain fault geometry and even some aspects of the rupture process at depth.

Orthogonal faulting occurring over a large region, at multiple length scales, implies that this is a primary characteristic of this fault zone. The smallest of these structures, having a length of 1 km or less, are probably being slowly generated as part of the ongoing formation of this young fault zone. Assuming a Mohr-Coulomb failure criterion, if two orthogonal planes are equally favorable, then the coefficient of friction must be close to zero, implying that the rocks are held together by cohesive strength alone (22). Alternatively, numerical studies have shown that cross-faults can be produced as a dynamic rupture effect near rupture edges (23). If a fault zone has widespread cross-faulting, the geometry could have a major influence on the earthquake rupture process, because at any point in time there will be many potential pathways to sustain the rupture propagation, thereby facilitating the occurrence of multifault ruptures.

The Ridgecrest sequence has many similarities with the 1987 Superstition Hills sequence (24), which also exhibited prominent orthogonal faulting and had an M_w 6.2 foreshock followed by a M_w 6.6 mainshock. In that sequence, the two largest events formed an L-shaped geometry, similar to the faults that ruptured during the Ridgecrest M_w 6.4 event. More generally, however, orthogonal faulting is present across much of southern California as well as Japan (25–28) and was also prominent in the 2012 Indian Ocean sequence (29).

The highly segmented nature of the Ridgecrest earthquakes suggests that the rupture process during these events could be more of a cascading phenomenon than a single continuous rupture front propagating along a fault. This behavior would depend at least partly on whether the faults are physically connected at depth. In a cascading rupture process, estimates of the average rupture velocity could be biased because the time of slip initiation for each segment is not a smoothly varying function of space and time. More important, however, is that the physics of the rupture process is entirely different in a cascading model, with ruptures potentially being modulated or driven by the seismic waves.

In the Ridgecrest sequence, at least 20 faults with a length of ~ 2 km or larger ruptured, with

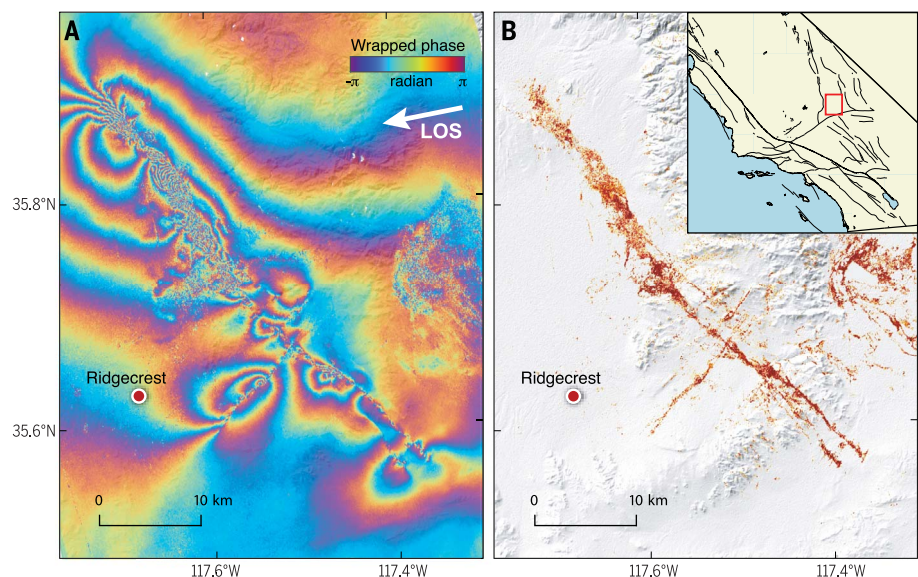


Fig. 3. Interferogram and DPM. (A) Coseismic interferogram derived from the ALOS-2 SAR image pair (2018-04-16 and 2019-07-08), showing the locations of surface ruptures. The line of site (LOS) is from ground to satellite. (B) DPM derived from coherence loss between pre- and postseismic Sentinel-1 SAR data. Darker colors indicate greater coherence losses.

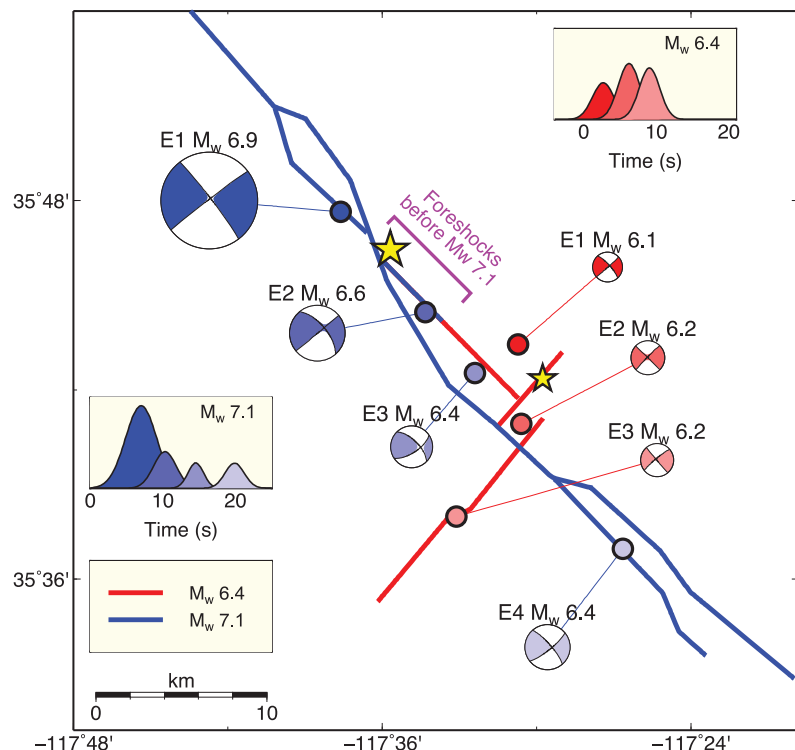


Fig. 4. Kinematic summary of rupture processes. The M_w 6.4 foreshock ruptured three main faults. This event was followed by foreshock activity along a northwest-trending fault and eventually triggered the M_w 7.1 event. The mainshock had four main subevents and ruptured bilaterally. Both events have very slow rupture velocities of ~ 2 km/s.

countless smaller faults that likely slipped too. Many of these ruptured the surface, while others are only visible at depth. The Ridgecrest earthquakes are the latest large crustal earthquakes to exhibit multifault ruptures (1–5), which

several decades ago were viewed as unlikely or outright impossible (1). Today, with the available resolution, this phenomenon appears to be more common than not. The persistence of multifault ruptures indicates that the

phenomenon is a critical element of the physics of earthquakes.

Furthermore, the regular occurrence of multi-fault ruptures has fundamental implications

for seismic hazard assessment. Accounting for all possible combinations of faults that could activate simultaneously during a single event or sequence is challenging, if not functionally

impossible. This task is made even more difficult by the fact that, as demonstrated by the Ridgecrest sequence, our database of large faults in California is still incomplete, and the concept

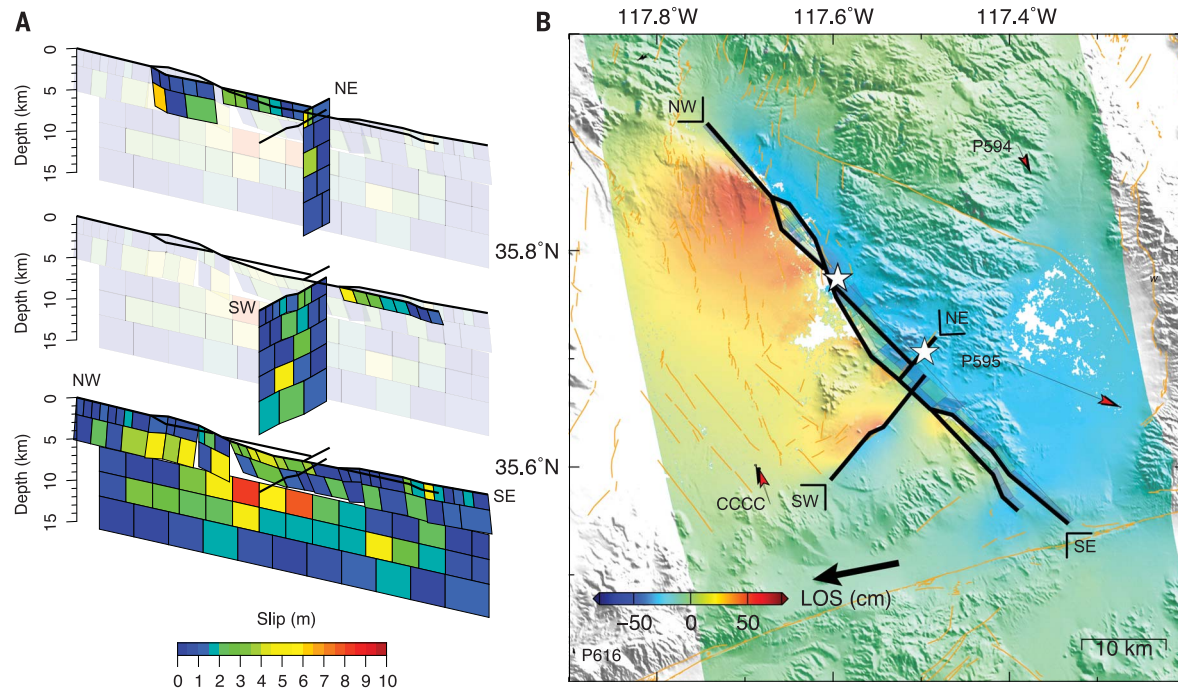
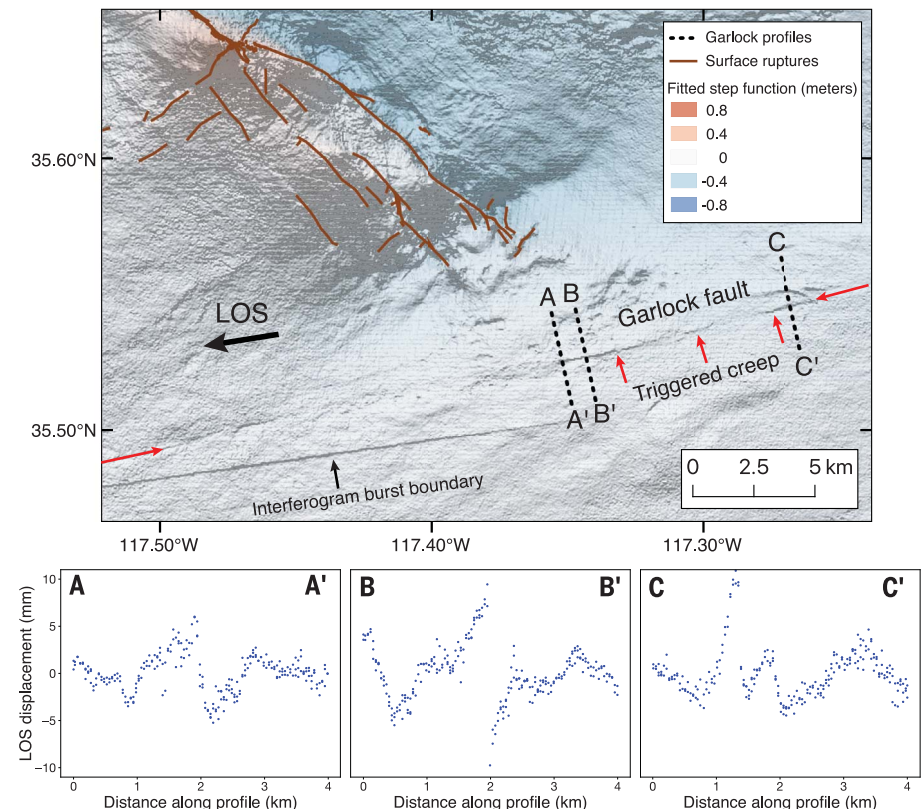


Fig. 5. Static inversion of geodetic data. (A) Mean slip model of the Ridgecrest sequence, including the slip contribution from the M_w 6.4 and M_w 7.1 events. (B) Black lines show the fault geometries used in the slip. Arrows indicate the data (black) and model prediction (red) of nearby GPS offsets. The color map indicates the displacements in the unwrapped ALOS2 ascending track 65 coseismic interferogram along the corresponding line-of-sight (LOS) direction (black arrow). Model predictions and data residuals are provided in the supplementary materials (II).

Fig. 6. Summary of triggered creep on the Garlock fault. Step function fit at the coseismic time for Sentinel 1, ascending track 64 InSAR time series, combined with gradient shading of the same field. The gradient shading reveals zones of surface offset along the Garlock fault. Fault perpendicular profiles reveal offset of up to 20 mm in the line-of-sight direction, while the width of the deformation profiles suggest that offset on the fault is confined to shallow depths.



of how faults are defined should be revisited. Magnitude estimates of historical earthquakes from paleoseismology could therefore be biased downward for many such multifault ruptures, while also exposing the potential for incorrectly splitting a single complex rupture into multiple smaller ones. As Fig. 1 shows, although the San Andreas fault may be the single largest source of hazard in southern California, nearly all large earthquakes since the 1906 San Francisco earthquake, over a century ago, have not occurred on the San Andreas fault, and many have been complex multifault ruptures.

Conclusions

The 2019 Ridgecrest sequence brought to an end the long earthquake silence in California. These events occurred within an immature fault zone and activated many orthogonal structures with lengths ranging from 1 km to more than 10 km. The largest events each ruptured multiple faults, a characteristic that has been repeatedly observed for large crustal earthquakes in recent years. Such scenarios are difficult to forecast for seismic hazard assessment. The rupture of the Ridgecrest mainshock terminated only a few kilometers from the Garlock fault, yet only aseismic creep was triggered at the closest section of the fault. Far to the southwest, a sizable swarm ensued, whereas triggered seismic activity on the entire eastern portion of the Garlock fault was negligible. At such close proximity to the mainshock rupture, the stress changes imparted by the mainshock are substantial. The last major earthquake occurred ~400 to 500 years ago (30). Future investigations that integrate the observed phenomena on the Garlock fault with geologic and geodetic observations will be important

for understanding its contribution to seismic hazard in the Eastern California shear zone.

REFERENCES AND NOTES

1. K. Sieh *et al.*, *Science* **260**, 171–176 (1993).
2. M. Simons, Y. Fialko, L. Rivera, *Bull. Seismol. Soc. Am.* **92**, 1390–1402 (2002).
3. D. Eberhart-Phillips *et al.*, *Science* **300**, 1113–1118 (2003).
4. S. J. Wei *et al.*, *Nat. Geosci.* **4**, 615–618 (2011).
5. I. J. Hamling *et al.*, *Science* **356**, eaam7194 (2017).
6. E. Hauksson *et al.*, *Seismol. Res. Lett.* **66**, 54–60 (1995).
7. C. B. Amos *et al.*, *GSA Bull.* **125**, 1187–1202 (2013).
8. G. A. Davis, B. C. Burchfiel, *Bull. Geol. Soc. Am.* **84**, 1407–1422 (1973).
9. Y. Fialko, M. Simons, *J. Geophys. Res. Solid Earth* **105** (B9), 21781–21793 (2000).
10. E. K. Haddon, C. B. Amos, O. Zielke, A. S. Jayko, R. Bürgmann, *Geochim. Geophys. Geosyst.* **17**, 2239–2269 (2016).
11. Materials and methods are described in the supplementary materials.
12. S.-H. Yun *et al.*, *Seismol. Res. Lett.* **86**, 1549–1556 (2015).
13. D. I. Doser, H. Kanamori, *J. Geophys. Res. Solid Earth* **91** (B1), 675–688 (1986).
14. E. Hauksson, *Geophys. J. Int.* **186**, 82–98 (2011).
15. N. Wetzler, T. Lay, E. E. Brodsky, H. Kanamori, *Sci. Adv.* **4**, eaao3225 (2018).
16. D. P. Hill *et al.*, *Science* **260**, 1617–1623 (1993).
17. Q. Zhang *et al.*, *Geophys. Res. Lett.* **44**, 726–733 (2017).
18. M. Wei, D. Sandwell, Y. Fialko, R. Bilham, *Geophys. Res. Lett.* **38**, n/a (2011).
19. J. F. Dolan, L. J. McAuliffe, E. J. Rhodes, S. F. McGill, R. Zinke, *Earth Planet. Sci. Lett.* **446**, 123–136 (2016).
20. X. Tong, D. T. Sandwell, B. Smith-Konter, *J. Geophys. Res. Solid Earth* **118**, 369–389 (2013).
21. Y. Ben-Zion, C. G. Sammis, *Pure Appl. Geophys.* **160**, 677–715 (2003).
22. C. H. Scholz, *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 2019).
23. S. Q. Xu, Y. Ben-Zion, *Geophys. J. Int.* **193**, 304–320 (2013).
24. K. W. Hudnut, L. Seeber, J. Pacheco, *Geophys. Res. Lett.* **16**, 199–202 (1989).
25. W. Thatcher, D. P. Hill, *Geology* **19**, 1116–1120 (1991).
26. E. Fukuyama, *Earth Planets Space* **67**, 38 (2015).
27. Z. E. Ross, E. Hauksson, Y. Ben-Zion, *Sci. Adv.* **3**, e1601946 (2017).
28. Z. E. Ross *et al.*, *Geophys. Res. Lett.* **44**, 8260–8267 (2017).
29. L. Meng *et al.*, *Science* **337**, 724–726 (2012).

30. C. Madden Madugo, J. F. Dolan, R. D. Hartleb, *Bull. Seismol. Soc. Am.* **102**, 2282–2299 (2012).
31. K. J. Kendrick *et al.*, *Southern California Earthquake Center Annual Meeting Proceedings* (SCEC, 2019), contribution 9779.
32. SCEC, Data are available from the Southern California Seismic Network (<https://doi.org/10.7914/SN/CI>), operated by the California Institute of Technology and the United States Geological Survey (2013).

ACKNOWLEDGMENTS

The ALOS-2 original data are copyrighted by JAXA and provided under JAXA RA6 PI projects P3278 and P3360. The Copernicus Sentinel-1 data were provided by the European Space Agency (ESA) and contain modified Copernicus data 2019, processed by ESA, Caltech, and NASA/JPL. **Funding:** E.H. received support from USGS/NEHRP grant G19AP00035 and NSF awards EAR-1550704 and EAR-1818582; Z.Z. and Z.J. received support from USGS grant G19AP00030. Part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This material is based on services provided by the GAGE Facility, operated by UNAVCO, Inc., with support from the National Science Foundation and the National Aeronautics and Space Administration under NSF Cooperative Agreement EAR-1724794. **Author contributions:** Z.E.R. performed the template matching and seismicity relocation. E.H. calculated three-dimensional earthquake locations. M.S. and B.I. performed the slip inversion. Z.J., X.W., and Z.Z. performed the subevent modeling. E.J.F. and O.L.S. performed the InSAR data processing. M.Z. determined offset fields for the InSAR data. S.-H.Y. and J.J. produced the DPMs. Z.L. and A.W.M. processed the GPS data. O.L.S. performed the Garlock fault creep analysis and smoothed seismicity plots. All authors contributed to writing the manuscript. **Competing interests:** A.W.M. is a member of the UNAVCO Geodetic Data Services Advisory Committee. **Data and materials availability:** All seismic data used are publicly available from the Southern California Earthquake Data Center and the IRIS DMC. GPS data are available from Network of the Americas (NOTA) and UNAVCO. The seismicity catalog produced in this study is publicly available from the Southern California Earthquake Data Center.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/346/suppl/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 to S4
References (33–61)

5 August 2019; accepted 23 September 2019
10.1126/science.aaz0109

GENE EXPRESSION

Genetic regulatory variation in populations informs transcriptome analysis in rare disease

Pejman Mohammadi^{1,2,3,4*}, Stephane E. Castel^{1,2}, Beryl B. Cummings^{5,6}, Jonah Einson^{1,2}, Christina Sousa^{3,4}, Paul Hoffman^{1,2}, Sandra Donkervoort⁷, Zhuoxun Jiang⁸, Payam Mohassel⁷, A. Reghan Foley⁷, Heather E. Wheeler^{9,10}, Hae Kyung Im⁸, Carsten G. Bonnemann⁷, Daniel G. MacArthur^{5,6}, Tuuli Lappalainen^{1,2*}

Transcriptome data can facilitate the interpretation of the effects of rare genetic variants. Here, we introduce ANEVA (analysis of expression variation) to quantify genetic variation in gene dosage from allelic expression (AE) data in a population. Application of ANEVA to the Genotype-Tissues Expression (GTEx) data showed that this variance estimate is robust and correlated with selective constraint in a gene. Using these variance estimates in a dosage outlier test (ANEVA-DOT) applied to AE data from 70 Mendelian muscular disease patients showed accuracy in detecting genes with pathogenic variants in previously resolved cases and led to one confirmed and several potential new diagnoses. Using our reference estimates from GTEx data, ANEVA-DOT can be incorporated in rare disease diagnostic pipelines to use RNA-sequencing data more effectively.

Large reference databases of human exomes and genomes have enabled the characterization of genomic variation in human populations (1–3). These data have been used to summarize genic intolerance to damaging variants, where depletion of gene-disrupting variants (e.g., stop-gain variants) indicates deleterious fitness consequences (1, 4, 5). Such analyses are essential for prioritizing rare and de novo coding variants that can underlie Mendelian disease and provide a genetic diagnosis for 25 to 50% of patients (6, 7). However, despite advances in DNA sequencing, the search for rare disease-causing variants outside the coding sequence has been hindered by the difficulty of interpreting rare regulatory variants and identifying their target genes.

Integration of genome- and transcriptome-sequencing data has provided improved diagnosis by better detection of rare variants with functional effects (6, 8–10). However, the often laborious analysis is further complicated by the transcriptome being affected by the environment, disease state, and technical variation. This has made it challenging to quantify when an effect is genetic and beyond the normal population range. Thus, most analyses have

been limited to only a small fraction of variants that induce clear alterations in the transcriptome, such as total loss of expression and splice defects.

One promising data type is the allelic expression (AE), which measures the relative expression of the paternal and maternal haplotype of a gene in an individual. Departure from equal AE, allelic imbalance, is largely unaffected by environmental and technical factors with a reported heritability of 85% (11), and therefore has a higher sensitivity to capture cis-acting genetic effects, including those induced by rare variants (6, 12–15). However, a quantitative framework for interpreting this data type to identify rare pathogenic variants has been lacking.

Here, we quantify the effects of genetic regulatory variation in populations using a mechanistic model of cis-regulatory variation. Specifically, for each gene, we estimate V^G , the expected variance in the dosage that is due to interindividual genetic differences within a population. Next, we use V^G as a reference to identify genes affected by potentially pathogenic regulatory variants in patients.

Results

Generative model for population allelic expression data and the ANEVA method

Cis-regulatory variant effect sizes can be quantified with allelic fold change (aFC) (16). aFC has an analytical link to gene dosage, which would allow calculation of V^G if all regulatory variants were known (supplementary materials, Eqs. 1 to 7). In practice, we can use AE data to estimate the overall distribution of regulatory effects on a gene without having to identify these variants explicitly. Across individuals, AE data represent a series of comparisons between the net expression effects of all variants on two random haplotypes at a time. A major complication for applications of AE data is

that, within a population, genes have diverse patterns depending on the properties of regulatory variants present and the single-nucleotide polymorphism (SNP) used to measure the allelic expression (aeSNP; Fig. 1, A to D) (14, 15). We derived a generative model for population AE data under a realistic scenario in which a gene is regulated by several regulatory variants, of which only some are identifiable. Under this assumption, population AE data are described by a constrained mixture of binomial-logit-normal (BLN) probability distribution functions (supplementary materials, Eqs. 8 to 19). We fit this model to population AE data (supplementary materials, Eqs. 20 to 28, and Fig. 1, E to H) and use the maximum likelihood parameters to estimate V^G indirectly (supplementary materials, Eqs. 29 and 30). We refer to this method as analysis of expression variation (ANEVA). Simulations show that the inferred V^G is accurate ($R^2 = 0.92$; fig. S1). Thus, ANEVA allows biologically interpretable estimates of genetic variation in gene expression within a population to be derived from AE read count data.

ANEVA estimates from AE data are consistent with eQTL data and heritability of gene expression

We applied ANEVA to 10,361 RNA-sequencing (RNA-seq) samples from 48 tissues and 620 individuals with whole-genome sequencing (WGS) data from the Genotype-Tissues Expression (GTEx) version 7 data (17, 18). Overall, we estimated V^G at a median of 43,219 autosomal aeSNPs per tissue. Gene-level V^G was derived as a weighted harmonic mean of SNP-level estimates for a median of 4962 genes per tissue and a total of 14,084 genes (Fig. 2, A and B, and table S1). First, we ensured that our AE-derived estimates were consistent with what is expected from expression quantitative trait locus (eQTL) data [median correlation = 0.73; Fig. 2C, fig. S2, and table S2]. Next, we benchmarked ANEVA estimates against gene expression cis heritability (h^2). For GTEx whole blood, we calculated the ratio of AE- and eQTL-derived V^G to the total variance of gene expression. These ANEVA-based h^2 estimates were consistent and comparable to those from studies using standard methods and with larger datasets, confirming that V^G measures the genetic variation in gene expression (Fig. 2D and fig. S3). Because AE-based ANEVA V^G estimates are more applicable to AE-based outlier detection, we used these estimates for all subsequent analyses (fig. S14) (19).

Genetically driven variation in gene expression across tissues, populations, and gene sets

Next, we analyzed how V^G varies between tissues and populations. The estimates were well correlated between tissues (median correlation = 0.57; Fig. 3A). For a given gene, V^G tends to be smaller in tissues where the gene is

¹New York Genome Center, New York, NY, USA. ²Department of Systems Biology, Columbia University, New York, NY, USA.

³Scripps Research Translational Institute, La Jolla, CA, USA.

⁴Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA.

⁵Analytical and Translation Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. ⁶Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁷Neuromuscular and Neurogenetic Disorders of Childhood Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA. ⁸Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA.

⁹Department of Biology, Loyola University Chicago, Chicago, IL, USA. ¹⁰Department of Computer Science, Loyola University Chicago, Chicago, IL, USA.

*Corresponding author. Email: pejman@scripps.edu (P. Mohammadi); tlappalainen@nygenome.org (T.L.)

more highly expressed (Wilcoxon signed-rank test $p < 10^{-300}$; Fig. 3B). Because this was not an artifact of differences in read depth (fig. S4), it suggests that there is an increased dosage sensitivity and a higher selective constraint in tissues where the gene has a more pronounced functional role (see fig. S5 for an example). To

analyze population differences in V^G , we used ANEVA on AE data from three European and one African subpopulation from the GEUVADIS data (20). We found a high correlation between estimates from all subpopulations (correlation range: [0.75, 0.83]; fig. S6). This suggests that the total amount of genetic dosage variation is

not highly variable between populations, and approaches that aggregate genetic effects at the gene level may have better applicability across populations than analyses of individual variants.

To characterize differences in the amount of genetic regulatory variation between genes,

Fig. 1. Cis-regulatory variation, allelic expression, and ANEVA.

(A to D) Examples of allelic expression across individuals (dots) for four genes with a single aeSNP each. In (A), similar haplotype expression levels for the gene indicate little cis-regulatory variation. In (B) to (D), there is relatively more variation. In (C) and (D), there are distinct clusters driven by different haplotype combinations of a common, strong regulatory variant and the aeSNP, with strong linkage disequilibrium in (D). These examples illustrate the challenge of consistently modeling the underlying regulatory variants. Ref., reference; Alt., alternative. (E) Schematic representation of ANEVA, which uses a generative model of population AE data and a mechanistic model of cis-regulatory variation to estimate the magnitude of genetic variation in expression for each gene. FC, fold change. (F to H) Generative model of population AE data represented mechanistically (F), in population AE data (G), and as a Bayesian plate diagram (H) (supplementary materials, Eqs. 20 to 22). AE data are modeled with one distinctly strong regulatory biallelic variant. If present, this variant is specified by its effect size, $s_{H,L}$, and a measure of linkage disequilibrium (LD) with the aeSNP. Residual cis-regulatory variation is modeled as an infinite-allelic regulatory variant summarized by variance term σ_r^2 . Allelic expressions e_R and e_A are measured at a heterozygous aeSNP with reference (R) and alternative (A) alleles, and $s_{R,A}^*$ is the aeSNP reference allele alignment bias. Haplotypes h_1 and h_2 , basal expression level e_B , and N cis-regulatory variant sites $v_1 \dots v_N$, are components of our complete formal model of cis-regulatory variation.

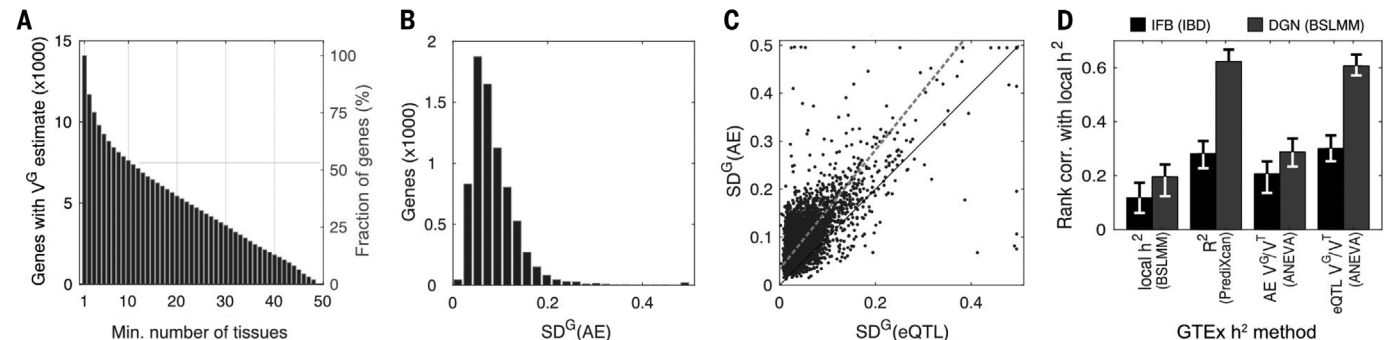
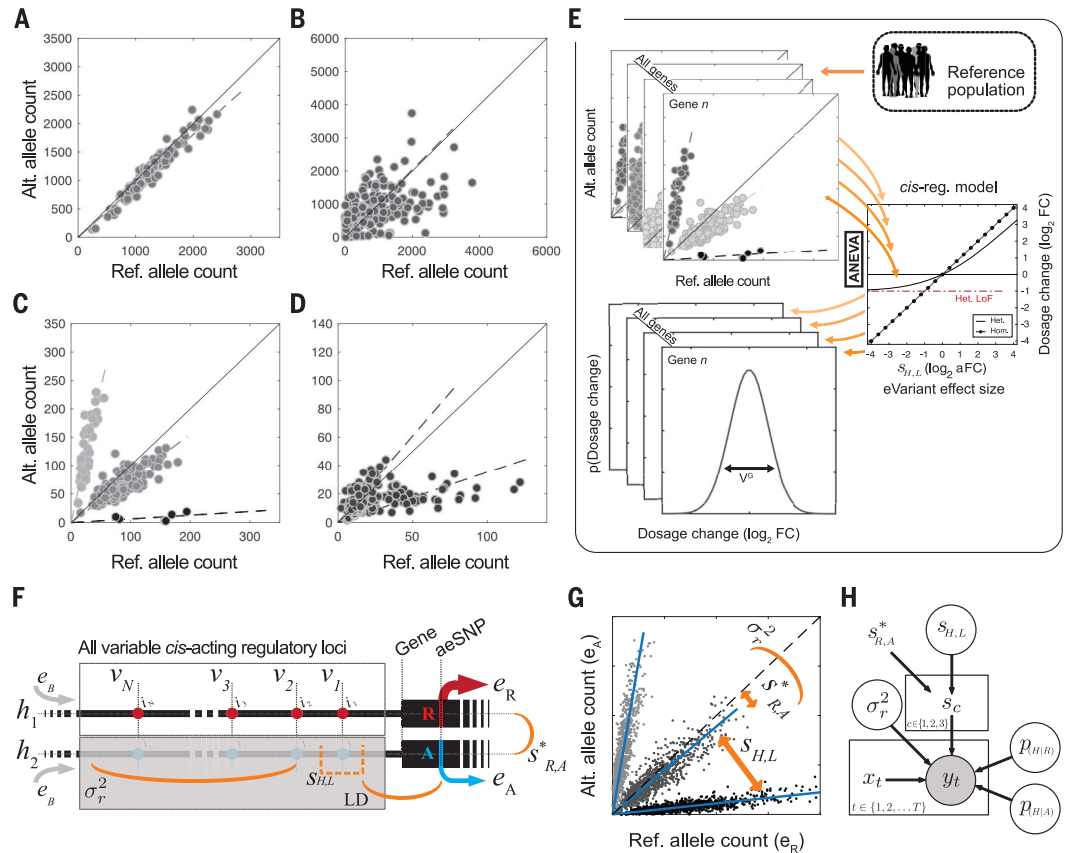


Fig. 2. Estimates of genetic regulatory variation in GTEx. (A) Number of genes with V^G estimates across one to 49 GTEx tissues. Min., minimum. (B and C) Distribution of SD^G , $\sqrt{V^G}$, for 7556 genes in GTEx subcutaneous adipose (B) and its comparison with eQTL data [(C); correlation (corr.) = 0.71]. The red line is Deming regression fit (fig. S2). SD^G is capped at 0.5 for visualization. (D) Benchmarking of ANEVA by gene expression heritability (h^2) estimates. GTEx h^2 was calculated by the linear mixed-model based BSLMM, PrediXcan R^2 , and ANEVA (19). These were compared with two larger cohorts: BLSMM h^2 from the Depression Genes and Networks cohort (DGN) cohort [$n = 922$ (32)] and local identity-by-descent (IBD)-based h^2 from the Iceland family blood (IFB) cohort [$n = 722$ (33)].

we correlated V^G to statistics of gene regulation and constraint. For each gene, we calculated a weighted harmonic mean of V^G across tissues (\bar{V}^G ; table S1). Gene enhancer size had a minimal correlation to \bar{V}^G (Fig. 3C) (21), suggesting that the size of the mutational target, a proxy for the background mutation rate, plays a minor role. Genes with high purifying selection for coding gene-disrupting variants or non-coding variants in the promoter or untranslated regions (UTRs) were depleted of genetic regulatory variation (Fig. 3C), as previously observed by eQTL analysis (1). Rare disease genes had lower \bar{V}^G , whereas loss-of-function tolerant genes had higher \bar{V}^G (Fig. 3D), showing that dosage sensitivity is captured by both exome and regulatory variation analysis. Genes identified by genome-wide association studies (GWAS) showed little deviation from the background, but schizophrenia genes had the lowest V^G and blood metabolite genes the highest, suggesting a link to genetic architecture of these traits. The amount of genetic regulation variation measured as V^G can complement previous coding and regulatory variation analyses of selective constraint on genes and traits.

Genetically driven variation in gene expression and dosage outlier testing from AE data

In addition to these biological insights, V^G has a direct practical application in identifying population outliers that may be pathogenic. To this end, we developed the ANEVA dosage outlier test (ANEVA-DOT) to identify genes

likely affected by a heterozygous genetic variant with an unusually strong effect on gene dosage. Using V^G for each gene, ANEVA-DOT tests against the null hypothesis that the observed allelic imbalance in an individual is consistent with dosage variation in the general population (Fig. 4A) while accounting for a number of additional technical and biological sources of variation (supplementary materials, Eqs. 31 to 42). We used extensive simulations to ensure that the test is well calibrated (fig. S7). ANEVA-DOT is implemented in R and runs in a few seconds per sample (22).

We first tested ANEVA-DOT in the general population of 466 skeletal muscle samples from GTEx. Each sample had a median of 3390 genes tested and 10 genes identified as outliers at a 5% false discovery rate (FDR) (hereafter referred to as ANEVA-DOT genes; 90% range: [3, 22]). An average of 56% of the genes previously implicated in neuromuscular disorders (6, 23) and up to 46% of the highly expressed genes were testable per individual (fig. S8). As a quality filter, 113 out of 5848 tested genes that appeared as outliers in >1% of the individuals were excluded from further analysis (fig. S8, D to F, and table S4). After this step, a median of 4.5 ANEVA-DOT genes were retained per individual (90% range: [1, 14]; Fig. 4B). ANEVA-DOT genes were highly enriched for rare heterozygous variants in a 10-kb window upstream of the transcriptional start site and in the gene body (Fig. 4B). This enrichment was particularly pronounced for rare

putative gene-disrupting variants that are expected to have a strong effect on gene expression levels by nonsense-mediated decay (Fig. 4, B and C). This confirms that ANEVA-DOT captures rare genetic effects on gene dosage.

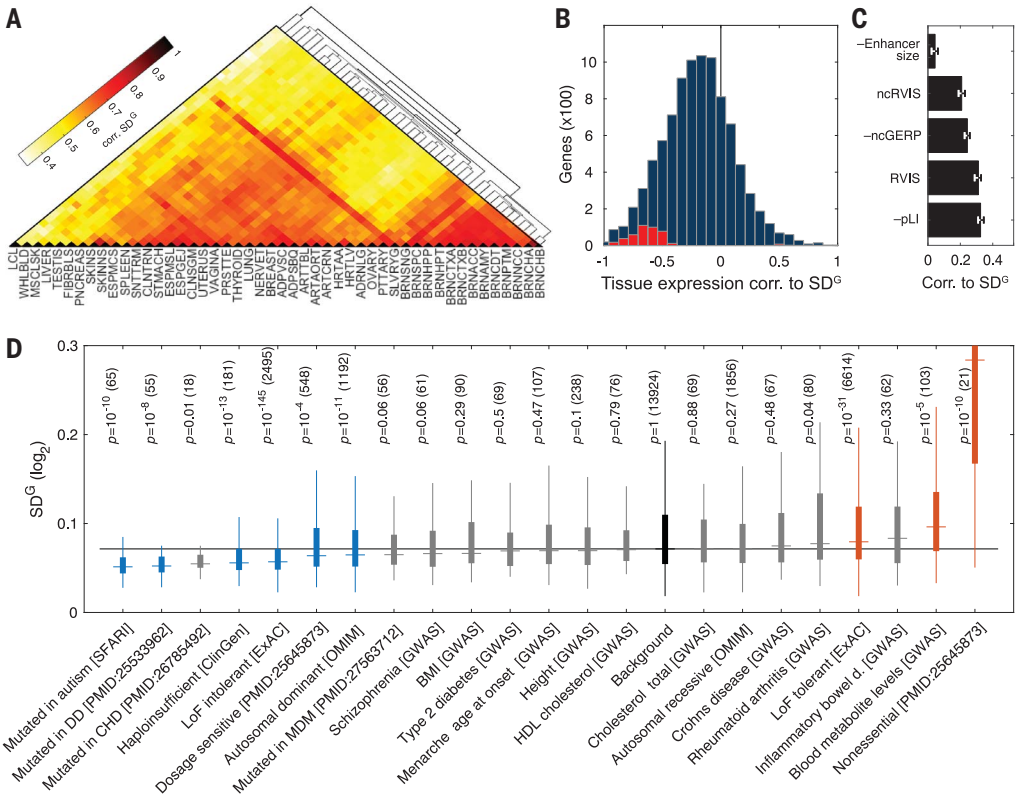
Next, we evaluated how sensitive ANEVA-DOT is to differences in the reference population where V^G is calculated. First, using the GEUVADIS data (20), we looked for ANEVA-DOT genes in 86 European (GBR) individuals using V^G estimates derived from two European populations (FIN and TSI) and one African population (YRI). The three reference populations performed similarly, with an average of 74% (range 69 to 78%) of ANEVA-DOT genes identified using one confirmed by another (fig. S9), suggesting that the lack of full concordance is likely driven by noise and threshold effects. However, larger sample sizes will be needed for a comprehensive evaluation of the population effects. Next, we investigated whether ANEVA-DOT genes in GTEx skeletal muscle could be identified by analyzing other accessible tissues of these individuals. The detection rate varied from 23.3% in fibroblast to 12.3% in whole blood, which indicates that ANEVA-DOT can capture some outlier effects also from proxy tissues (fig. S10).

ANEVA-DOT accurately identifies disease genes in AE data from rare disease patients

To test ANEVA-DOT's performance in the diagnosis of rare disease patients, we applied it to AE data from 70 rare Mendelian muscle

Fig. 3. Biological sources of regulatory variation between genes.

(A) Correlation (corr.) of genetic regulatory variation across GTEx tissues (see table S1 for tissue names). (B) Rank correlation between median expression in a tissue and V^G for 9158 genes with V^G estimates in at least five tissues. The distribution is shifted (median rank correlation -0.20). Significant genes are shown in red (5% FDR). (C) Rank correlation of V^G with enhancer size, coding constraint (RVIS, pLI), noncoding constraint (ncRVIS), and noncoding conservation (ncGERP) in UTRs and promoters. (D) V^G for different gene sets (DD, developmental disorder; CHD, congenital heart disease; MDM, congenital muscular dystrophies and myopathies; table S3), with nominal p -values from rank-sum test compared with the background of all genes ($p \leq 0.01$ highlighted) and the number of genes in parentheses. Boxes span the middle 50% values and the whiskers span ± 1.5 interquartile range (IQR) from first and the third quartile.



dystrophy and myopathy (MDM) patients using the V^G reference from GTEx skeletal muscle (figs. S11 to S17, and table S5). Of the 65 patients with high-quality data, 32 had a previous diagnosis, of which 21 were expected to lead to allelic imbalance (6). These cases were used as positive controls to benchmark ANEVA-DOT against previous tests of allelic imbalance: binomial and beta-binomial tests, binomial test with an allelic imbalance threshold, and a naïve population-aware test of excess allelic imbalance against GTEx data using the z -test (Fig. 4, D to H, and fig. S12). ANEVA-DOT identified a median of 11 outlier genes per individual (out of a median of 2190 tested), substantially fewer than other tests, (Fig. 4H).

This small number of outliers always included the previously diagnosed gene when there was a detectable allelic imbalance present (76%; figs. S11 and S12), typically (69%) among the top five most significant genes (table S5). ANEVA-DOT's high recall and precision outperformed all of the other tests by a substantial margin (Fig. 4I and figs. S12 to S14) (19).

In the 33 patients without a genetic diagnosis from previous whole-exome sequencing (WES) and/or WGS or RNA-seq analysis (6), we found a median of nine ANEVA-DOT genes per sample (in total 349 genes), which included at least one neuromuscular disease gene (6, 23) in 12 patients (in total 17 genes; figs. S15 and S16). One of these potential new diagnoses

from ANEVA-DOT was confirmed: Patient N10, with a limb-girdle muscular dystrophy-like phenotype, had 13 ANEVA-DOT genes, with the one known Mendelian muscle disease gene, *DES*, being the most statistically significant. Further RNA-seq and reverse transcription polymerase chain reaction analysis identified a pseudo-exon insertion caused by a variant creating an intronic splice site. This had been missed by the prior gene panel, WES, WGS, and RNA-seq analysis because of challenging in silico interpretation of intronic variants and the relatively low number of RNA-seq reads. The variant is in trans with a pathogenic missense variant that had not been identified as a diagnosis because of the lack of a second

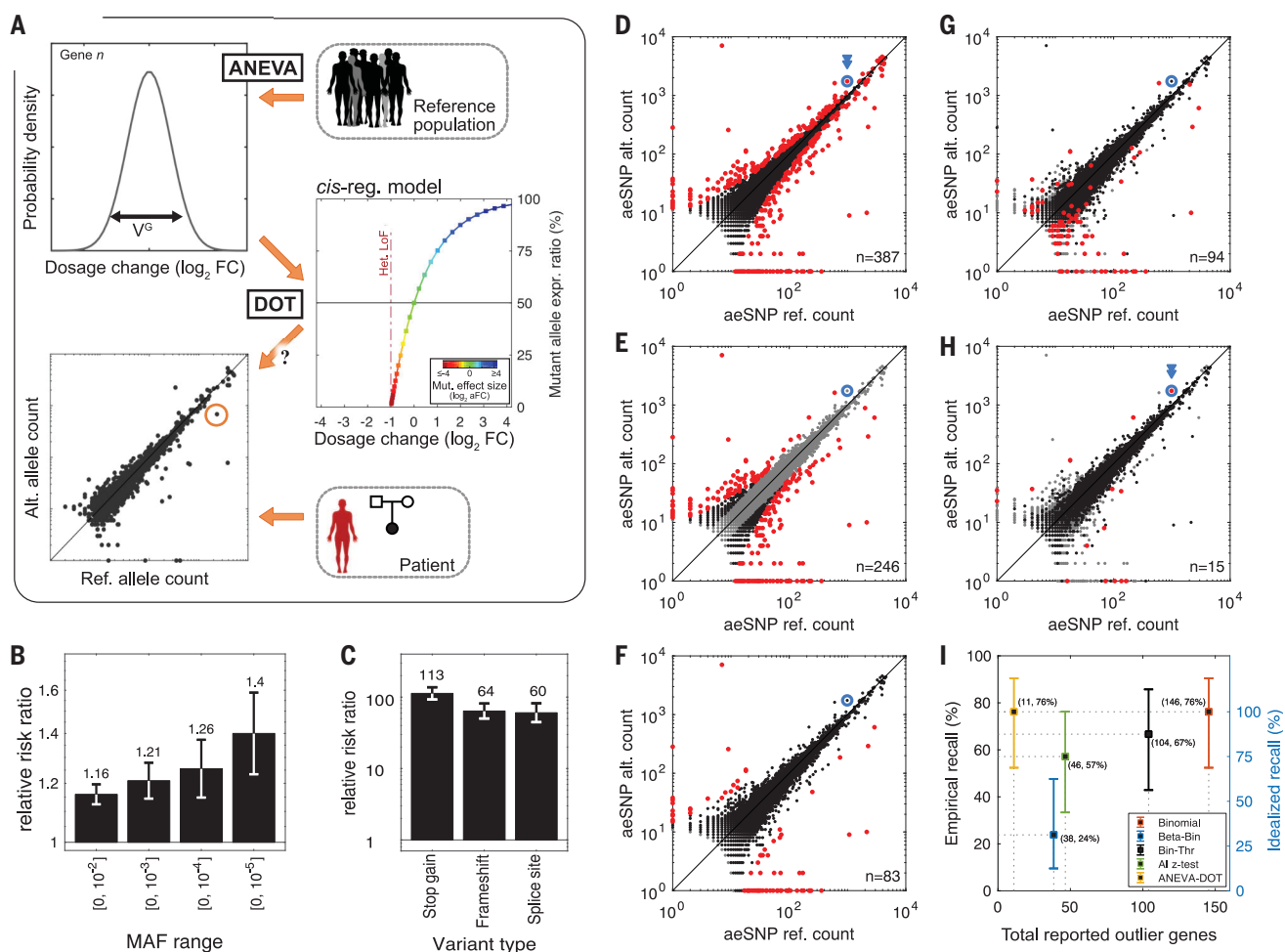


Fig. 4. Regulatory outliers detected with the ANEVA-DOT. (A) Illustration of the ANEVA-DOT method. For each gene, the null distribution of allelic imbalance is estimated using the V^G and the model of cis-regulatory genetic effect. Allelic counts in a test individual are compared with this null, accounting for sampling noise, sequencing noise, reference bias, and the variant haplotype. (B and C) Enrichment of all rare variants in ANEVA-DOT genes as a function of allele frequency (B) and for putative gene-disrupting variants [minor allele frequency < 1% (C)]. (D to H) Example of AE data for all genes from one previously diagnosed muscle dystrophy patient (N13). The disease gene is shown in blue. Outlier genes identified by different tests (5% FDR) are marked in red.

binomial [$n = 387$ (D)], binomial with a 15% allelic imbalance threshold [Bin-Thr, $n = 246$ (E)], beta-binomial [Beta-Bin, $n = 83$ (F)], excess allelic imbalance against GTEx data by z -test [AI z -test, $n = 94$ (G)], and ANEVA-DOT [$n = 15$ (H)]. Genes marked in gray are excluded from each test. (I) Fraction of true causal genes identified in previously diagnosed patients (recall) and its 95% bootstrap confidence intervals versus the number of outliers reported. Empirical recall (left) is calculated using all cases in which imbalanced AE would be expected ($n = 21$), whereas idealized recall (right) excludes five cases in which detecting the gene from AE data is impossible (e.g., when the causal gene is not expressed; fig. S12).

variant (fig. S18). Additionally, ANEVA-DOT identified strong candidates in six cases and possible candidates in 11 others (19). By design, ANEVA-DOT does not rely on identifying which variant underlies the dosage-outlier effect, but genetic analysis can be applied after prioritizing genes by ANEVA-DOT. This is currently mostly limited to gene- or splice-disrupting variants because of their easier annotation compared with rare regulatory variant candidates that may also exist. Overall, we expect up to 10.5 of the 17 known MDM and 18.8 of all 349 identified ANEVA-DOT genes in the 33 undiagnosed patients to be true disrupted causative genes (19).

Discussion

In this study, we introduce a method, ANEVA, and its extension, ANEVA-DOT, to quantify genetic variation in gene dosage in the general population, and to identify genes where a patient appears to carry a heterozygous variant with an unusually strong effect on gene expression. This enables individual transcriptome comparison to previously generated reference data without the caveats of technical and reverse causation noise in total gene expression analysis.

The ANEVA framework uses biologically interpretable units of gene dosage, allowing interpretation of regulatory and coding gene-disrupting variants on the same scale. Furthermore, the statistical methods introduced here for modeling allelic expression data are applicable to other uses of this data type.

ANEVA-DOT is a fast and powerful approach for finding genes with likely disease effects, with the small numbers of outliers making further manual curation feasible in a clinical setting without compromising sensitivity. The use of V^G estimates from GTEx as a shared reference for ANEVA-DOT analysis of patients is analogous to the use of coding constraint metrics for prioritization of pathogenic coding variants. ANEVA-DOT outlier genes can be further prioritized by candidate gene lists and by tools that are currently used in exome sequencing follow-up (1, 2, 5, 24, 25). Because ANEVA-DOT captures transcriptome outcomes of genetic effects without having to identify rare regulatory variants themselves, this method is particularly advantageous for rare genetic effects from poorly defined reg-

ulatory elements, but it will also detect, for example, variants triggering transcript decay. However, identifying the specific variants underlying ANEVA-DOT outliers is still challenging despite existing variant prioritization approaches, especially for noncoding regions (26–28).

Despite these advantages, our methods have several limitations. The main caveat is that AE data are sparse, and V^G estimates may be lacking or noisy for genes with few common coding variants owing to small size, high coding constraint, or low expression levels. These issues will, however, improve with increasingly large RNA-seq datasets. ANEVA-DOT is only applicable to about half of expressed genes per individual that have an aeSNP. Finally, allelic imbalance is not informative of recessive effects without family analysis. Thus, like other genetic diagnosis tools, ANEVA-DOT should be used in conjunction with other methods to capture different types of rare variants underlying disease. We envision that in clinical genetics, when practically feasible, transcriptome data will become a powerful additional layer of data for interpreting the genome and its disease-contributing variants.

REFERENCES AND NOTES

- M. Lek et al., *Nature* **536**, 285–291 (2016).
- K. J. Karczewski et al., *bioRxiv* 531210 (2019); <https://doi.org/10.1101/531210>.
- 1000 Genomes Project Consortium, *Nature* **526**, 68–74 (2015).
- J. M. Havrilla, B. S. Pedersen, R. M. Layer, A. R. Quinlan, *Nat. Genet.* **51**, 88–95 (2019).
- S. Petrovski, Q. Wang, E. L. Heinzen, A. S. Allen, D. B. Goldstein, *PLOS Genet.* **9**, e1003709 (2013).
- B. B. Cummings et al., *Sci. Transl. Med.* **9**, eaal5209 (2017).
- L. Farnæs et al., *NPJ Genom. Med.* **3**, 10 (2018).
- L. S. Kremer et al., *Nat. Commun.* **8**, 15824 (2017).
- H. D. Gonorazky et al., *Am. J. Hum. Genet.* **104**, 466–483 (2019).
- L. Frésard et al., *Nat. Med.* **25**, 911–919 (2019).
- A. Buil et al., *Nat. Genet.* **47**, 88–91 (2015).
- X. Li et al., *Nature* **550**, 239–243 (2017).
- X. Wang, D. B. Goldstein, *bioRxiv* (2018).
- D. M. McKean et al., *Nat. Commun.* **7**, 12824 (2016).
- E. C. Glassberg, Z. Gao, A. Harpak, X. Lan, J. K. Pritchard, *Genetics* **211**, 757–772 (2019).
- P. Mohammadi, S. E. Castel, A. A. Brown, T. Lappalainen, *Genome Res.* **27**, 1872–1884 (2017).
- GTEx Consortium et al., *Nat. Genet.* **45**, 580–585 (2013).
- GTEx Consortium, *Nature* **550**, 204–213 (2017).
- See the supplementary materials.
- T. Lappalainen et al., *Nature* **501**, 506–511 (2013).
- S. Fishilevich et al., GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, (2017).
- P. Mohammadi, C. Sousa, ANEVA-DOT software package (2019); <https://doi.org/10.5281/zenodo.3406690>.
- J.-C. Kaplan, D. Hamroun, *Neuromuscul. Disord.* **25**, 991–1020 (2015).
- J. Birgmeier et al., AMELIE accelerates Mendelian patient diagnosis directly from the primary literature. *bioRxiv* 1171322 (2017); <https://doi.org/10.1101/171322>.
- P. Deelen et al., *Nat. Commun.* **10**, 2837 (2019).
- P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, *Nucleic Acids Res.* **47** (D1), D886–D894 (2019).
- J. di Iulio et al., *Nat. Genet.* **50**, 333–337 (2018).
- D. Anderson, T. Lassmann, *NPJ Genom. Med.* **3**, 5 (2018).
- P. Mohammadi, P. Hoffman, R package for BLN distribution functions (2019); <https://doi.org/10.5281/zenodo.3406692>.
- P. Mohammadi, ANEVA software package (2019); <https://doi.org/10.5281/zenodo.3406688>.
- P. Mohammadi, PejLab/Datasets: GTEx V7 Vg and freq estimates (2019); <https://doi.org/10.5281/zenodo.3406717>.
- GTEx Consortium, A. Battle et al., *Genome Res.* **24**, 14–24 (2014).
- A. L. Price et al., *PLOS Genet.* **7**, e1001317 (2011).

ACKNOWLEDGMENTS

We thank the GTEx consortium, especially F. Aguet, N. Ferraro, and S. Montgomery. **Funding:** This work was funded by the following NIH grants: UL1TR002550 (P. Mohammadi), UL1TR001114 (P. Mohammadi), K99HG009916 (S.E.C.), R01MH106842 (T.L., P.H.), UM1HG008901 (T.L., J.E.), R01GM122924 (T.L., R01MH107666 (H.K.I. and H.E.W.), P30DK020595 (H.K.I.), R15HG009569 (H.E.W.), and UM1 HG008900 (D.G.M.), as well as intramural funds from the NIH (C.G.B., A.R.F., P. Mohassel, S.D.). **Author contributions:** P. Mohammadi and T.L. designed the study. P. Mohammadi developed all statistical models. P. Mohammadi, S.E.C., B.C., and J.E. analyzed the data. P. Mohammadi, C.S., and P.H. developed software tools. S.D., Z.J., P. Mohassel, A.R.F., H.E.W., H.K.I., C.G.B., and D.G.M. provided data and materials. P. Mohammadi and T.L. wrote the paper with input from all authors. **Competing interests:** D.G.M. is a founder with equity in and T.L. is an adviser for Goldfinch Bio. S.E.C. is a cofounder, chief technology officer, and stock owner for Variant Bio. T.L. is an adviser with equity in Variant Bio. H.K.I. has received honoraria from GSK and AbbVie. The other authors have no competing interests. **Data and materials availability:** Software packages for the presented methods are available online as follows: BLN distribution functions (29), ANEVA (30), and ANEVA-DOT (22). Outlier summary statistics for all GTEx tissues are available in (31). The GTEx v7 data and MDM cohort data are available in dbGaP (phs000424.v7.p2 and phs000655.v3.p1, respectively). Gene-level AE data and ANEVA-DOT results in the MDM cohort are provided in supplementary data S1.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/351/suppl/DC1
Material and Methods
Supplementary Text
Tables S1 to S5
Data S1
Figs. S1 to S18
References (34–47)

[View/request a protocol for this paper from Bio-protocol.](#)

13 May 2019; accepted 24 September 2019
10.1126/science.aay0256

REPORT

EXOPLANETS

Oxygen fugacities of extrasolar rocks: Evidence for an Earth-like geochemistry of exoplanets

Alexandra E. Doyle^{1*}, Edward D. Young^{1*}, Beth Klein², Ben Zuckerman², Hilke E. Schlichting^{1,2,3}

Oxygen fugacity is a measure of rock oxidation that influences planetary structure and evolution. Most rocky bodies in the Solar System formed at oxygen fugacities approximately five orders of magnitude higher than a hydrogen-rich gas of solar composition. It is unclear whether this oxidation of rocks in the Solar System is typical among other planetary systems. We exploit the elemental abundances observed in six white dwarfs polluted by the accretion of rocky bodies to determine the fraction of oxidized iron in those extrasolar rocky bodies and therefore their oxygen fugacities. The results are consistent with the oxygen fugacities of Earth, Mars, and typical asteroids in the Solar System, suggesting that at least some rocky exoplanets are geophysically and geochemically similar to Earth.

Estimating the composition of extrasolar planets from host-star abundances or from planet mass-radius relationships is difficult and unreliable (1, 2). The elemental abundances in some white dwarfs (WDs) provide an alternative, more direct approach for determining the composition of extrasolar rocks. WDs are the remnant cores left behind when a star ejects its hydrogen-rich outer layers after the red giant phase. These remnant cores are $\sim 0.5 M_{\odot}$ (solar masses) and about the same radius as Earth, are no longer powered by fusion, and slowly cool over time. Because of their high densities, and thus strong gravitational fields, elements heavier than helium rapidly sink below their surfaces, becoming unobservable. Nonetheless, spectroscopic studies show that the atmospheres of up to half of WDs with effective temperatures $< 25,000$ K are “polluted” by elements heavier than He (3–5). The source of these heavy elements is exogenous, coming from accretion of debris from rocky bodies that previously orbited the WDs (6–9). We exploit this pollution to measure the elemental constituents of extrasolar rocky bodies. We collated observations from the literature of polluting elements in six WDs: SDSS J104341.53+085558.2 (10), SDSS J122859.92+104033.0 (9), SBSS 1536+520 (11), GD 40 (8, 12), SDSS J073842.56+183509.6 (13), and LBQS 1145+0145 (14) (hereafter, SDSS J1043+0855, WD 1226+110, WD 1536+520, GD 40, SDSS J0738+1835, and WD 1145+017, respectively). Their coordinates are listed in table S1. The bulk compositions of the bodies polluting these WDs resemble those of rocky bodies in the Solar System (15, 16) (Fig. 1).

We use the relative abundances of rock-forming elements in polluted WDs to determine the effective partial pressure of oxygen, i.e., the oxygen fugacity (f_{O_2}) of the accreted rocks. Oxygen fugacity is a measure of the degree of oxidation in the rocks. It corresponds to the effective partial pressure of gaseous oxygen that would be in thermodynamic equilibrium with

the material of interest. In combination with other factors, the intrinsic oxygen fugacity of a planet will determine the relative size of its metallic core, the geochemistry of its mantle and crust, the composition of its atmosphere, and the forces responsible for mountain building (17, 18). Oxygen fugacity is also thought to be among the parameters that determine the habitability of a planet (19). In practice, f_{O_2} is usually expressed as the nonideal partial pressure of oxygen relative to a convenient reference value.

Oxygen fugacities of rocky planets are often reported relative to the reference Iron-Wüstite (IW) equilibrium reaction $Fe(Fe) + \frac{1}{2} O_2 = FeO$ (Wüstite), such that $\Delta IW \equiv \log(f_{O_2}) - \log(f_{O_2})_{IW}$ (16). When expressed this way, differences in oxygen fugacity are nearly independent of temperature and pressure (16). The initial oxidation state of a rocky body with at least some Fe metal at the time of its formation is recorded by the concentration of oxidized iron (hereafter denoted as FeO, although it may include other oxides of iron) in the rock and the concentration of Fe in the metal

$$\Delta IW = 2 \log \left(\frac{x_{FeO}^{rock}}{x_{Fe}^{metal}} \right) + 2 \log \left(\frac{\gamma_{FeO}^{rock}}{\gamma_{Fe}^{metal}} \right) \quad (1)$$

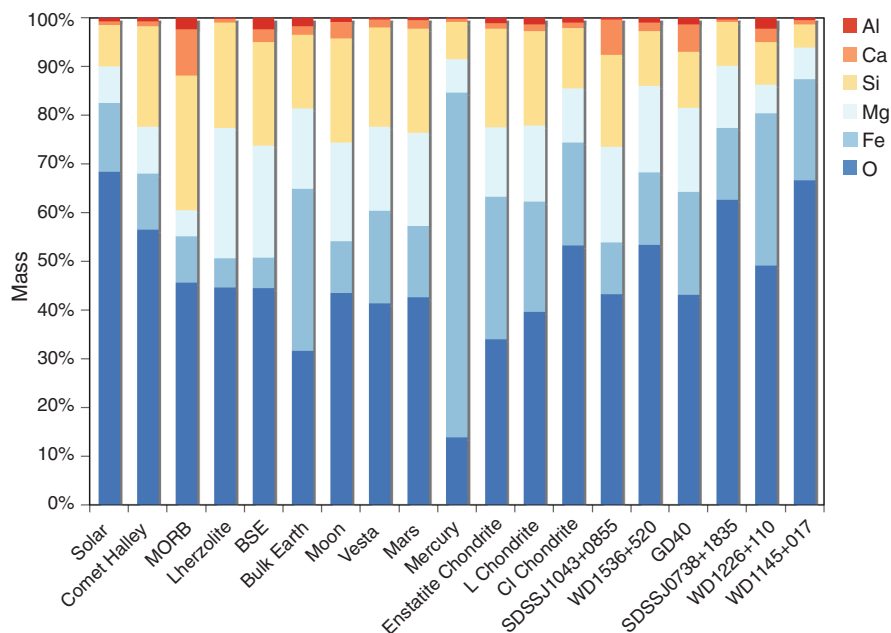


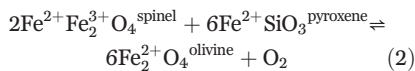
Fig. 1. Bulk compositions by mass for six white dwarfs compared with Solar System bodies. Bulk compositions of the six rock-forming elements Al, Ca, Si, Mg, Fe, and O are indicated by the colored bars. The six white dwarfs are shown in the right-most columns. Shown for comparison are Solar System objects: the Sun, Comet Halley (1P/Halley), Earth, the Moon, Vesta, Mars, Mercury, three types of meteorites (enstatite chondrite, ordinary L chondrite, and carbonaceous CI chondrite), and three terrestrial igneous rock types [mid ocean ridge basalt (MORB), lherzolite (representing Earth's mantle), and bulk silicate Earth (BSE)] (16). The relatively high abundances of Fe in bulk Mercury and bulk Earth are due to their metal cores. The compositions of the white dwarfs are similar to the Solar System rocks. The large amount of O in WD 1145+017 is highly uncertain. Values are listed in data S1.

¹Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, Los Angeles, CA, USA.

²Department of Physics and Astronomy, University of California, Los Angeles, Los Angeles, CA, USA. ³Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA.

*Corresponding author. Email: a.doyle@ucla.edu (A.E.D.); eyoung@epss.ucla.edu (E.D.Y.)

where x_i^k are mole fractions of the species i in phase k , γ_i^k are activity coefficients for the species, and thermodynamic activities are $a_i^k = x_i^k \gamma_i^k$. To facilitate comparison, we set the uncertain activity coefficients to unity, so the second term on the right-hand side of Eq. 1 vanishes. Equation 1 expresses the f_{O_2} at the time the planet or planetesimal formed (20, 21); we refer to this as the intrinsic oxygen fugacity of the body. The partitioning of iron between rock and metal during formation leaves a record of the intrinsic oxygen fugacity in the form of the mole fraction of FeO in the rocks, x_{FeO}^{rock} . This signature persists even after the rock and metal are separated by the process of differentiation (partitioning between core and mantle). This is because changes in the valence state of iron during subsequent reactions proceed without appreciably altering the total amount of iron bonded to oxygen in the rocks. For example, the reaction



determines the Fe^{3+}/Fe^{2+} ratio, and thus the f_{O_2} , in a rock containing the minerals spinel, pyroxene, and olivine, without substantially altering the total Fe bonded to oxygen (from 2.17 oxygens per Fe to 2.00 oxygens per Fe). Reactions like these that follow the formation of a rocky body lead to local variations in f_{O_2} within the body but do not generally alter the intrinsic oxygen fugacity recorded by application of Eq. 1 (Fig. 2). The intrinsic oxygen fugacity of Earth is constrained by $x_{FeO}^{mantle} = 0.06$ [8 wt % (weight percent) FeO] in its mantle and the composition of its Fe-rich core. This leads to a terrestrial ΔIW value of about -1 to -2 , with the range due to the uncertain values for the activity coefficient ratio [commonly used values of $\gamma_{FeO}^{mantle}/\gamma_{Fe}^{core}$ range from ~ 1 to 4 (22)].

The material accreted by the six polluted WDs in this study are rocks devoid of metal, as demonstrated by the lack of excess Fe relative to oxygen (Fig. 1). Separation of metal and rock during accretion onto WDs is suggested by the ranges in element ratios in polluted WDs (23) and from observations of a metal-density planetesimal core orbiting a WD (24). Our f_{O_2} measurements are representative of the fugacity values at the time of core formation, even though they are derived from crustal or mantle rocks. The maximum intrinsic oxygen fugacity calculated from Eq. 1 approaches 0 as the mole fraction of Fe in the silicate increases. Values for ΔIW greater than ~ -0.9 for elemental concentrations similar to Solar System rocks imply that all of the iron has been oxidized, and that the intrinsic ΔIW values from Eq. 1 are therefore minimum estimates for the oxidation state at the time the rocks formed.

The solar protoplanetary disk must, on average, have had the same composition as the Sun (see supplementary text in the supplementary materials). The oxygen fugacity of a gas with solar composition is determined by its H_2O/H_2 ratio, after correcting for the oxygen bound to carbon in CO and other less abundant oxides, according to the reaction $H_2 + \frac{1}{2} O_2 \rightleftharpoons H_2O$ (16). Studies of meteorites reveal that, like Earth, most rocky bodies in the Solar System formed with ΔIW approximately five orders of magnitude higher than that of a solar gas (25, 26) (Fig. 3). The presence of large amounts of iron bonded to oxygen in silicates in chondrite meteorites indicates there was a relatively high oxygen fugacity during the earliest stages of rock formation in the Solar System (27). The enhancement in oxygen fugacity during rocky body formation may be attributable to the sublimation of water-rich and/or rock-rich dust at high dust/gas ratios (28). In this context, we examine whether the processes that led to oxidation of rocks in the Solar System are typical of other planetary systems, and therefore whether the geophysical and geochemical characteristics of Earth are likely to be common among rocky exoplanets.

When the six major rock-forming elements are measured in a polluted WD, the abundance

of FeO may be used to determine the oxidation state of the accreted exoplanetary rocky bodies (2, 8, 29). Polluted WDs with observed abundances of O, Mg, Si, Fe, Al, and Ca can be used to calculate oxygen fugacities from Eq. 1 by recognizing that any Fe not bonded to oxygen must have existed as metal in the accreted bodies. Data for polluted WDs are preferable to elemental abundances in other stars because of the known rocky provenance of the accreted elements in WDs, especially in the case of oxygen.

Our basic methodology is as follows: The oxide components MgO, SiO₂, FeO, Al₂O₃, and CaO describe the compositions of the major minerals that make up the accreting rocks. By assigning oxygen first to Mg, then Si, Al, Ca, and finally Fe, we calculate the relative amount of oxidized Fe, as FeO, and assign any remaining Fe to metal representing the core of the body (2, 8, 30). We propagate measurement uncertainties for the polluted WDs using a Monte Carlo bootstrap approach (16).

We validated our method using Solar System bodies by converting the composition of these bodies into hypothetical polluted WDs, as if rocks from the bodies (e.g., Earth, Mars, Mercury) had accreted onto a WD. We used typical WD measurement uncertainties for these calculations and recovered the known intrinsic

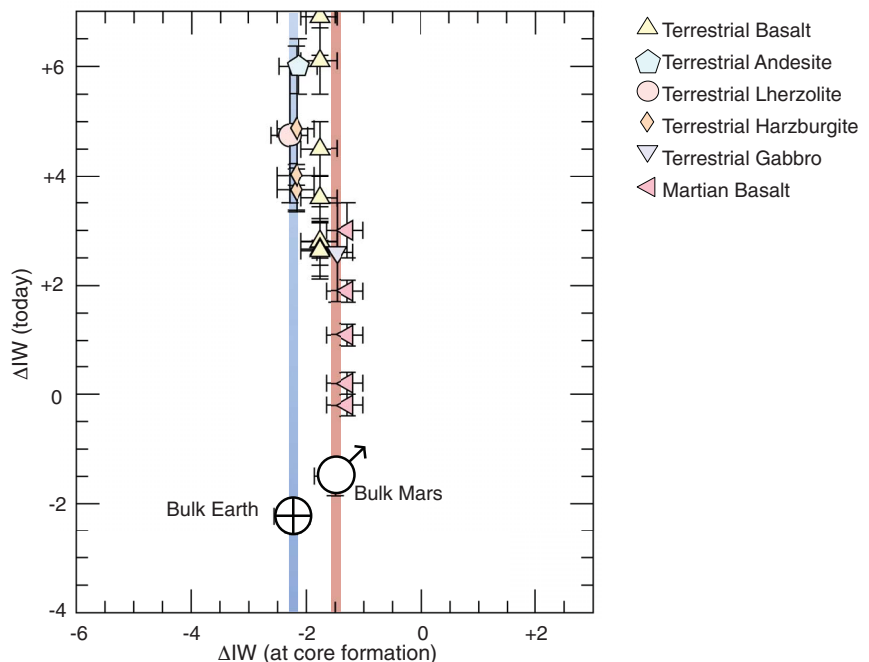


Fig. 2. Oxygen fugacities relative to IW at core formation versus today. Terrestrial and Martian rocks are characterized by ΔIW at the time of core formation, as calculated from the concentration of FeO, and ΔIW as measured today using various other measures of oxygen fugacity (16). Bulk Earth and bulk Mars values are also shown, demonstrating their similarity in intrinsic ΔIW at the time of core formation, despite the ranges in ΔIW as measured today. Error bars are 1σ (16). Where oxygen fugacities were previously reported relative to the quartz-fayalite-magnetite buffer (QFM), we converted them using $\Delta IW = \Delta QFM - 4$. Andesite, basalt, and gabbro represent crustal rocks, whereas lherzolite and harzburgite, specific types of peridotites, are mantle rocks.

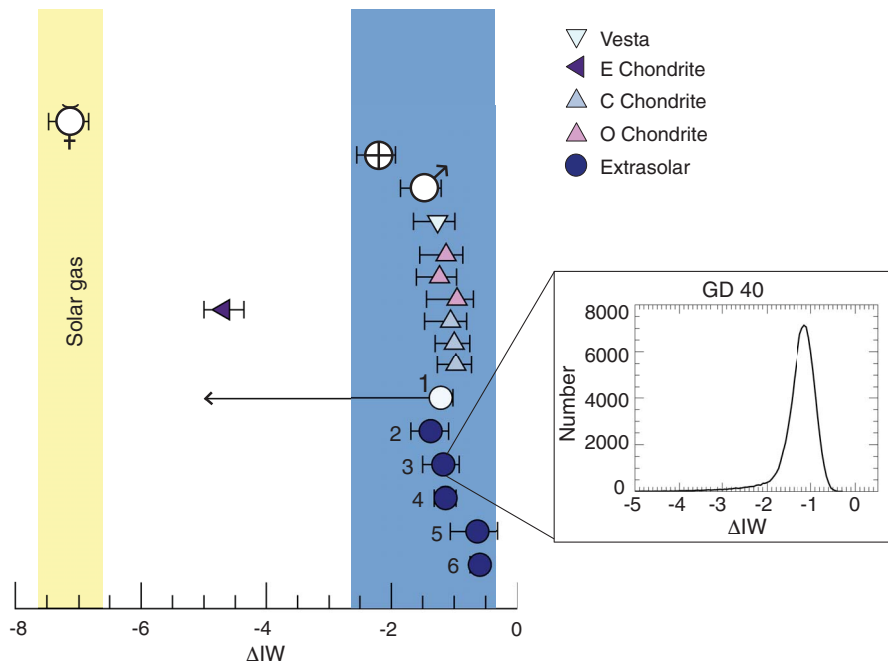


Fig. 3. Calculated oxygen fugacities relative to IW for rocky extrasolar bodies. Numbered circles show the values for rocky debris that polluted the white dwarfs: (1) SDSS J1043+0855; (2) WD 1536+520; (3) GD 40; (4) SDSS J0738+1835; (5) WD 1226+110; and (6) WD 1145+017. Values are listed in table S1. Error bars (1σ) are from propagation of measurement uncertainties (16). Only an upper limit could be obtained for SDSS J1043+0855 owing to the measurement uncertainties relative to the Fe concentration for that star (16). The ranges of relative oxygen fugacities for a gas of solar composition (yellow) and for most Solar System rocky bodies (blue) are shown for comparison. Rocks from Solar System planets are also shown and are represented by their planet symbols: Earth (\oplus), Mars (\mars), and Mercury (\mercury). Triangles show values for meteorites, representing bodies in the asteroid belt, including Vesta. The inset shows an example ΔIW probability distribution for one of the WDs, GD 40 (16); equivalents for the other WDs are shown in fig. S3.

oxygen fugacities for Earth, Mars, Mercury, Vesta, and various chondritic bodies (16). The Solar System bodies span a range in ΔIW of ~ 6 dex, in agreement with previous studies showing that Mercury and enstatite meteorites have f_{O_2} orders of magnitude lower than those for Earth, Mars, and other chondrite group meteorites (31, 32).

The six WDs in this study were chosen because quantitative measurements of all six major rock-forming elements are available for each. These WDs also exhibit infrared excesses, indicative of surrounding debris disks (6). The ΔIW values we obtain for the rocks accreted by the polluted WDs are all similar to those of Earth, Mars, Vesta, and the asteroids represented by carbonaceous (C) and ordinary (O) chondrites in the Solar System (Fig. 3).

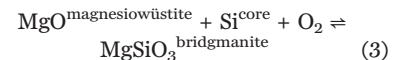
In cases where there is both more oxygen than required to oxidize all other major elements and a commensurate amount of hydrogen, water in the accreting body is implied and the partitioning of oxygen between ice and Fe can be ambiguous (33, 34). In addition, unaccounted-for Si in metal cores may have liberated oxygen to oxidize Fe in the accreted bodies. We find that

these effects are small for the WDs in this study and do not affect the elemental abundances we used to derive oxygen fugacities (16) (fig. S5).

The high oxygen fugacities of these extrasolar rocks, relative to a solar gas, suggest that whatever process oxidized rock-forming materials in the Solar System also operated in these other planetary systems. The large amount of oxidized iron in chondrite meteorites shows that oxidation relative to a solar gas occurred early in the Solar System, evidently before, or during the earliest stages of, planetesimal formation. Raising ΔIW by 5 log units, from solar to rock-like values, requires the gas to acquire an H_2O/H_2 ratio ~ 400 times that of a solar gas (28). This enrichment factor is greater than can be explained by simply transporting water in the form of ice particles from the outer to the inner Solar System (28). If dust/gas ratios control the oxidation states during rock formation, we conclude that the Solar System and the planetary systems around these six polluted WDs had similar ratios. This implies that high dust/gas ratios are intrinsic to rock formation in protoplanetary disks. A similar composi-

tional link between planet formation in the Solar System and that around other stars is indicated by the depletion of carbon in both solar and extrasolar rocks (35, 36).

The high oxidation state of these rocks determined the mineralogy, and therefore the geophysical behavior, of their parent bodies or the planets these bodies formed. For example, the lower mantle of Earth is composed of $\sim 70\%$ bridgmanite and $\sim 20\%$ magnesio-wüstite, two mineral phases with markedly different rheological properties whose abundances depend on f_{O_2} . The relative abundances of these minerals determine the dynamic behavior of the mantle (37). For illustration, the influence of oxygen fugacity on the mineralogy of a silicate mantle, and the composition of a metal-rich core, can be described by the reaction



where MgO refers to the Mg component in mantle magnesio-wüstite $[(\text{Mg,Fe})\text{O}]$, MgSiO_3 refers to the Mg component (bridgmanite) in mantle silicate perovskite $[(\text{Mg,Fe})\text{SiO}_3]$, and Si^{core} refers to Si in the metal-rich core. Rearranging the equilibrium constant for the reaction in Eq. 3, $k_{\text{Eq. 2}}$, shows that the activity ratio of bridgmanite to magnesio-wüstite is expected to vary with oxygen fugacity

$$f_{O_2} = \frac{a_{\text{MgSiO}_3}^{\text{bridgmanite}}}{a_{\text{MgO}}^{\text{magnesio-wüstite}} a_{\text{Si}}^{\text{core}} k_{\text{Eq. 2}}} \quad (4)$$

Equation 4 also illustrates that the Si content of the core varies inversely with oxygen fugacity. The concentrations of Si and other light elements in the core likely play a role in driving the compositional convection within the core that powers Earth's magnetic field (38, 39), which affects a planet's habitability (40). The relative size of the metallic core of a body (or even its existence) is also determined by oxygen fugacity (41). If the body or bodies that accreted onto WD 1536+520 were otherwise similar to Earth or its antecedents, the ΔIW value of -1.37 would result in a planet with an Fe-rich metal core making up $\sim 20\%$ of the mass of the parent body. For comparison, the most highly oxidized bodies we found, with $\Delta IW \sim -0.6$, would assemble to form a planet with no Fe-rich metal core.

Our results show that the parent objects that polluted these WDs had intrinsic oxidation states similar to those of rocks in the Solar System. Based on estimates of their mass, the bodies accreting onto WDs were either asteroids that represent the building blocks of rocky exoplanets, or they were fragments of rocky exoplanets themselves (15, 42). In either case,

our results constrain the intrinsic oxygen fugacities of rocky bodies that orbited the progenitor star of their host WD. Our data indicate that rocky exoplanets constructed from these planetesimals should be geophysically and geochemically similar to rocky planets in the Solar System, including Earth.

REFERENCES AND NOTES

1. A. Gupta, H. E. Schlichting, *Mon. Not. R. Astron. Soc.* **487**, 24–33 (2019).
2. C. T. Unterborn, W. R. Panero, *Astrophys. J.* **845**, 61 (2017).
3. B. Zuckerman, D. Koester, I. N. Reid, M. Hunsch, *Astrophys. J.* **596**, 477–495 (2003).
4. B. Zuckerman, C. Melis, B. Klein, D. Koester, M. Jura, *Astrophys. J.* **722**, 725–736 (2010).
5. D. Koester, B. T. Gänsicke, J. Farihi, *Astron. Astrophys.* **566**, A34 (2014).
6. M. Jura, *Astrophys. J.* **584**, L91–L94 (2003).
7. B. Zuckerman, D. Koester, C. Melis, B. M. Hansen, M. Jura, *Astrophys. J.* **671**, 872–877 (2007).
8. B. Klein, M. Jura, D. Koester, B. Zuckerman, C. Melis, *Astrophys. J.* **709**, 950–962 (2010).
9. B. T. Gänsicke, D. et al., *Mon. Not. R. Astron. Soc.* **424**, 333–347 (2012).
10. C. Melis, P. Dufour, *Astrophys. J.* **834**, 1 (2016).
11. J. Farihi et al., *Mon. Not. R. Astron. Soc.* **463**, 3186–3192 (2016).
12. M. Jura, S. Xu, B. Klein, D. Koester, B. Zuckerman, *Astrophys. J.* **750**, 69 (2012).
13. P. Dufour et al., *Astrophys. J.* **749**, 6 (2012).
14. S. Xu, M. Jura, P. Dufour, B. Zuckerman, *Astrophys. J.* **816**, L22 (2016).
15. M. Jura, E. D. Young, *Annu. Rev. Earth Planet. Sci.* **42**, 45–67 (2014).
16. Materials and methods are available as supplementary materials.
17. D. T. Wetzel, M. J. Rutherford, S. D. Jacobsen, E. H. Hauri, A. E. Saal, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8010–8013 (2013).
18. K. Putirka, *Am. Mineral.* **101**, 819–840 (2016).
19. B. L. Ehlmann et al., *J. Geophys. Res. Planets* **121**, 1927–1961 (2016).
20. J. Siebert, J. Badro, D. Antonangeli, F. J. Ryerson, *Science* **339**, 1194–1197 (2013).
21. D. C. Rubie, C. K. Gessmann, D. J. Frost, *Nature* **429**, 58–61 (2004).
22. M. K. Jordan, H. Tang, I. E. Kohl, E. D. Young, *Geochim. Cosmochim. Acta* **246**, 461–477 (2019).
23. M. Jura, E. D. Young, *Annu. Rev. Earth Planet. Sci.* **42**, 45–67 (2014).
24. C. J. Manser et al., *Science* **364**, 66–69 (2019).
25. J. I. Simon et al., *Earth Planet. Sci. Lett.* **238**, 272–283 (2005).
26. L. Grossman, J. R. Beckett, A. V. Fedkin, S. B. Simon, F. J. Ciesla, *Rev. Mineral. Geochem.* **68**, 93–140 (2008).
27. L. Grossman, A. V. Fedkin, S. B. Simon, *Meteorit. Planet. Sci.* **47**, 2160–2169 (2012).
28. J. Monteux, G. J. Golabek, D. C. Rubie, G. Tobie, E. D. Young, *Space Sci. Rev.* **214**, 39 (2018).
29. D. C. Rubie et al., *Earth Planet. Sci. Lett.* **301**, 31–42 (2011).
30. M. Asplund, N. Grevesse, A. J. Sauval, *Nucl. Phys. A* **777**, 1–4 (2006).
31. L. R. Nittler, S. Z. Weider, *Elements* **15**, 33–38 (2019).
32. C. Cartier, B. J. Wood, *Elements* **15**, 39–45 (2019).
33. J. Farihi, B. T. Gänsicke, D. Koester, *Science* **342**, 218–220 (2013).
34. R. Raddi et al., *Mon. Not. R. Astron. Soc.* **450**, 2083–2093 (2015).
35. M. Jura, *Astrophys. J.* **653**, 613–620 (2006).
36. J.-E. Lee, E. A. Bergin, H. Nomura, *Astrophys. J.* **710**, L21–L25 (2010).
37. J. Girard, G. Amulele, R. Farla, A. Mohiuddin, S. Karato, *Science* **351**, 144–147 (2016).
38. G. Helffrich, *Earth Planet. Sci. Lett.* **391**, 256–262 (2014).
39. D. Gubbins, Q. J. R. Astron. Soc. **32**, 69–84 (1991).
40. P. Cuartas-Restrepo, *Open Astron.* **27**, 183–231 (2018).
41. L. T. Elkins-Tanton, S. Seager, *Astrophys. J.* **688**, 628–635 (2008).
42. J. Farihi et al., *Mon. Not. R. Astron. Soc.* **424**, 464–471 (2012).

ACKNOWLEDGMENTS

This paper benefited from the comments provided by three anonymous reviewers. **Funding:** E.D.Y. acknowledges support from the NASA Exoplanets program grant NNX16AB53G. A.E.D. acknowledges financial support from NASA Space Grant. H.E.S. gratefully acknowledges support from the National Aeronautics and Space Administration under grant 17 NAI18_2-0029 issued through the NExSS Program. **Author contributions:** A.E.D. performed the calculations and co-wrote the manuscript. E.D.Y. conceived of the project, supervised the calculations, and co-wrote the manuscript. B.Z. and B.K. analyzed the WD data. H.E.S. contributed to the statistical analysis. All authors contributed to the interpretation of the data and the preparation of the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Input compositional data are available in data S1 and our derived oxygen fugacities are listed in table S1. The bootstrap and analytical data analysis codes are available as data S2 and S3.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/356/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Table S1
References (43–71)
Data S1 to S3

19 March 2019; accepted 20 September 2019
10.1126/science.aax3901

3D PRINTING

Rapid, large-volume, thermally controlled 3D printing using a mobile liquid interface

David A. Walker^{1,2*}, James L. Hedrick^{2,3*}, Chad A. Mirkin^{1,2,3†}

We report a stereolithographic three-dimensional printing approach for polymeric components that uses a mobile liquid interface (a fluorinated oil) to reduce the adhesive forces between the interface and the printed object, thereby allowing for a continuous and rapid print process, regardless of polymeric precursor. The bed area is not size-restricted by thermal limitations because the flowing oil enables direct cooling across the entire print area. Continuous vertical print rates exceeding 430 millimeters per hour with a volumetric throughput of 100 liters per hour have been demonstrated, and proof-of-concept structures made from hard plastics, ceramic precursors, and elastomers have been printed.

Additive manufacturing has been a disruptive force (1–7). Traditionally, it has been used for prototyping, substantially reducing the time from product ideation to manufacturing, but it has recently begun to make notable progress in the manufacturing sector (8, 9). One promising approach for the printing of plastic objects is stereolithography (SLA), which uses photocurable liquid resins that can be chemically formulated to yield materials with a large range of attractive mechanical properties (10, 11). Conventional SLA operates by using ultraviolet (UV) light to cure liquid photoactive resins on a vertically moving plate; the result is a desired three-dimensional (3D) object consisting of stacked 2D layers (12). DeSimone and co-workers introduced a variation of SLA called continuous liquid interface printing, which uses oxygen inhibition to create a reaction “dead layer” (13, 14). This dead layer prevents adhesion between the emerging part and the bottom of the print vat, obviating the need to repeatedly mechanically cleave the part from the vat (12). This continuous print approach increases vertical print speeds by two orders of magnitude while simultaneously removing material defects intrinsic to the aforementioned layer-by-layer lamination approaches. However, the polymerization reactions used in SLA photoreins are highly exothermic, and at such print speeds, the challenge of heat dissipation is daunting (15, 16). Moreover, the delivery of gaseous oxygen, a thermal insulator, through the print bed to create a dead layer limits one to peripheral cooling options that cannot rapidly dissipate the heat being generated.

We report a dead layer-free approach to rapid SLA printing, HARP (high-area rapid printing), which is capable of continuously

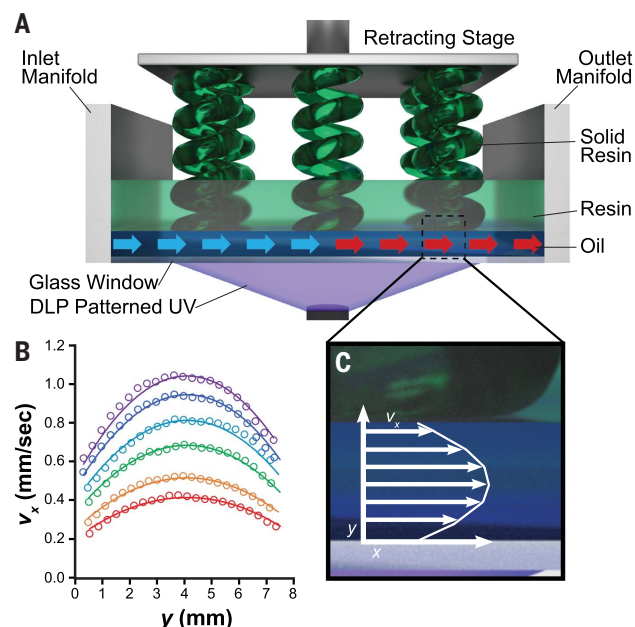
printing over large areas and at rapid vertical print speeds. The printer operates on the principle of a UV-curable resin floating on a bed of flowing immiscible fluorinated oil to minimize interfacial adhesion at the build region. Aizenberg and co-workers have demonstrated the great breadth of dewetting behavior possible when using fluorinated liquids embedded within solid matrices for applications in medicine and marine biofouling (17, 18). HARP builds on this concept by keeping the fluorinated phase in constant motion relative to the emerging printed part, further decreasing the adhesion forces (i.e., static versus dynamic) and generating a solid-liquid slip boundary. The oil can be recirculated through a heat exchanger to cool or heat the build region and maintain thermostatic control across the entire print bed. Additionally, the oil can be continuously filtered to remove the microparticulate

solids that are generated during the SLA print process and decrease resolution through light scattering (commonly referred to as “clouding”) (19). Finally, because HARP does not require an oxygen dead layer, it is compatible with both oxygen-sensitive and -insensitive ink chemistries, increasing the scope of applicable resins and resulting materials.

HARP operates through a mobile liquid interface that creates a shear stress beneath the emerging part and results in a slip boundary (Fig. 1). The slip boundary allows for the solidified part to be continuously retracted from the print interface. Fluorinated oils (perfluoropolyether copolymers, such as Solvay Fomblin Y or Chemours Krytox GPL) were chosen for their omniphobic properties and higher densities relative to that of common SLA resins. Although we explored other immiscible liquid systems, including densified water and glycerin, neither of these aqueous options produced the same quality of dewetting behavior with the emerging 3D printed part; both resulted in lower-quality print production (see fig. S7). A pair of inlet and outlet manifolds were fabricated to distribute a laminar flow of oil across the print bed with a uniform velocity profile (fig. S1). This flow profile ensures that the oil layer remains optically uniform across the build platform (i.e., no turbulence giving rise to optical distortions at the oil/resin interface) and results in a uniform interfacial-shear stress being applied to all solidified parts.

To confirm our hypothesis of a slip boundary being central to the HARP printer's operation, we used particle imaging velocimetry to analyze the cross-sectional fluid flow profile as

Fig. 1. Flow profile of a mobile interface that enables continuous printing. (A) Scheme of a 3D printed part emerging from the HARP 3D printer. (B) Velocity profile under printed part at different flow speeds, demonstrating the presence of a slip boundary. Colors represent increasing volumetric fluxes q (red, $q = 0.21$ mm/s; orange, $q = 0.30$ mm/s; green, $q = 0.44$ mm/s; teal, $q = 0.56$ mm/s; blue, $q = 0.66$ mm/s; violet, $q = 0.75$ mm/s). Open circles are experimental data points from particle-imaging velocimetry; continuous lines are fits from an analytical model. (C) Scheme inset of the slip boundary flow profile under the part, with a representative experimentally observed flow profile.



¹Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. ²International Institute for Nanotechnology, Northwestern University, Evanston, IL 60208, USA. ³Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: chadnano@northwestern.edu

the oil passed beneath a printed part (20). Slip boundary conditions are notoriously difficult to observe because of the randomized optical reflections that occur at the interfaces under study, resulting in noisy data (21, 22). To rectify this problem, we fit an analytical model, allowing for the possibility of either a slip or non-slip boundary condition, to the center of the flow profile (i.e., where high-quality data can be collected). The model allowed for a classical Navier slip boundary to arise when the interfacial shear stress exceeded a given critical shear stress (23). From this analysis, the slip boundary model best captured the dynamics of the experimental data (Fig. 1B and fig. S3) (24). This is best reflected in the experimental velocity profiles [$v_x(y)$; circles in Fig. 1B] as a function of the oil flow rate (increasing volumetric flux, from red to violet) and the corresponding Navier slip model fits (solid lines in Fig. 1B).

In the absence of heat removal, rapid vertical print speeds of even modestly sized parts quickly exceed the smoke point of the resin. These extreme temperature swings ($\Delta T \approx >120^\circ\text{C}$) result in variable reactivities across the print bed, thermal warping/cracking of the printed material, and uncontrolled clouding that limits lateral resolution; ultimately these factors result in a deformed object and failed print (fig. S6) (24). To quantify the ability of HARP to dissipate heat, we performed thermal imaging of a small test print (5 cm \times 5 cm cross-sectional dimensions) in a specially designed miniature print vat (i.e., low-profile walls to enable imaging across the bed). Three different flow scenarios were evaluated: (i) printing without a mobile interface or active cooling (Fig. 2A); (ii) printing with a mobile interface to actively dissipate heat, but with no cooling of the oil (Fig. 2B); and (iii) printing with a mobile interface and active cooling of the oil (Fig. 2C). Without flow to actively dissipate heat, the printed part quickly exceeded the sensor range of the infrared (IR) camera at 150°C , and exceeded the flashpoint of the resin's primary monomer diluent (1,6-hexanediol diacrylate, 113°C). Similar experiments involving the generation of parts having larger cross-sectional areas (e.g., 20 cm \times 20 cm) exceeded 180°C (as evidenced by an IR noncontact thermometer gun measurement) in the absence of an active flow to dissipate the accumulating heat. In the case presented here, the accumulation of heat (Fig. 2A) led to surface temperatures in excess of 150°C and the part cracked during the printing process, causing a physical displacement from the stage (observable in the final frame of Fig. 2A, indicated by white lines overlaid on the part; see also fig. S6). Alternatively, the flow of oil depicted in Fig. 2B dissipated this heat into an oil reservoir. Over time, the temperature of the reservoir increased, reducing its effectiveness. Last, in Fig. 2C the

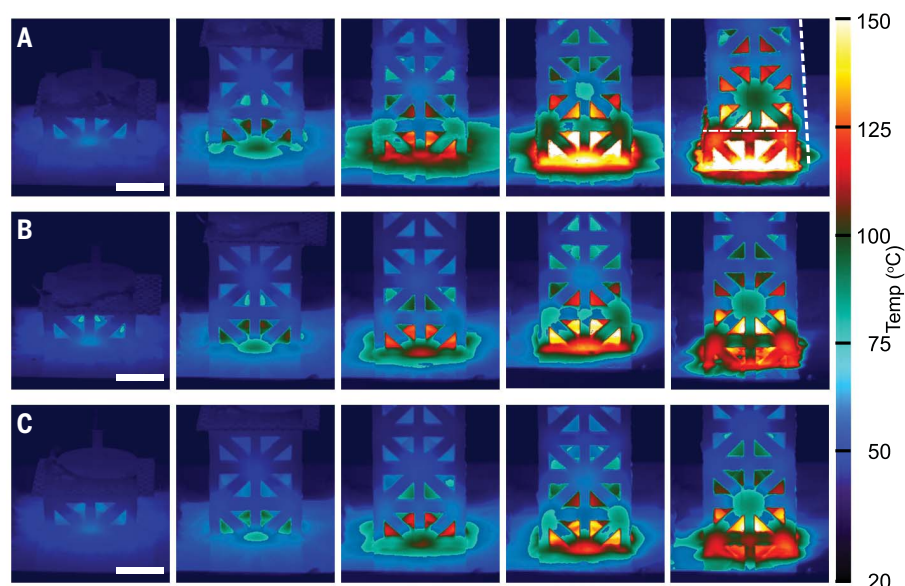


Fig. 2. IR thermal images of an emerging 3D printed part (hard polyurethane acrylate resin; cross section, 5 cm \times 5 cm; vertical print rate, 120 $\mu\text{m/s}$; optical resolution, 100 μm) under three different print conditions. (A) Stationary print interface. (B) Mobile interface. (C) Mobile interface with active cooling. Elapsed time between panels (left to right) is ~ 500 s; scale bars, 25 mm. Data and thermal color mapping correspond to movies S1 to S3.

oil was actively cooled while flowing to help stabilize the reservoir temperature, as well as the surface temperature of the emerging part, at 100° to 120°C .

To further evaluate the scalability of HARP through cooling, we printed a 38 cm \times 61 cm \times 76 cm part in 1 hour and 45 min with volumetric throughput of 100 liters/hour (fig. S4). At this scale, the fidelity of the print was affected by the resin's contractility [i.e., the volumetric reduction undergone during the resin solidification process, common to all SLA resins and different for each resin formulation (25)], because even modest contractions of a few percent were amplified into observable bowing. In contrast, when smaller single parts were produced, this contractility became less evident and print fidelity increased (Fig. 3C). Full use of the large-scale capabilities enabled by HARP will require the development of new SLA resins with lower volumetric contractility.

By not relying on oxygen quenching, HARP is capable of printing both oxygen-sensitive and oxygen-insensitive resin chemistries. To demonstrate this ability, we printed three different resins on the HARP 3D printer to generate three different classes of materials with varied properties: a hard polyurethane acrylate, an elastomeric butadiene rubber, and a silicon carbide ceramic (see Fig. 4). These resins are a mix of resins found in the literature (26, 27), obtained from companies, and formulations developed for the HARP printer. As with all

SLA systems, each resin must be optimized in terms of initiators and blockers to perform in conjunction with the light engine used and the desired vertical print rate. This was optimized by altering the photo-initiators used and tuning their concentrations to achieve a desired energy cure and penetration depth response for each resin (24). The hard polyurethane resin printed optimally under a vertical print speed of 120 $\mu\text{m/s}$. The ceramic resin used thiolene-click chemistry, a chemistry known to be relatively oxygen-insensitive (28, 29), to cross-link the polymeric resin and produce a "green" part. This polymeric silicone can be post-treated in a furnace at 1000°C to undergo pyrolysis to eliminate the organic components and produce a silicon carbide structure (26). Finally, the butadiene rubber resin required a slower print speed (30 $\mu\text{m/s}$) owing to its lower reactivity and higher viscosity. Although the postprocessing and print conditions differed for each resin (Fig. 4) (24), all the resins were compatible with the HARP approach. This makes HARP printing generally useful with stereolithographic resins currently available through literature or commercial formulations (26, 29).

To determine whether the HARP continuous print process produces isotropic materials, we performed mechanical testing on dog-bone structures printed from an ABS (acrylonitrile butadiene styrene)-like urethane acrylate resin in multiple orientations. This was done using the protocols of ASTM standard D638

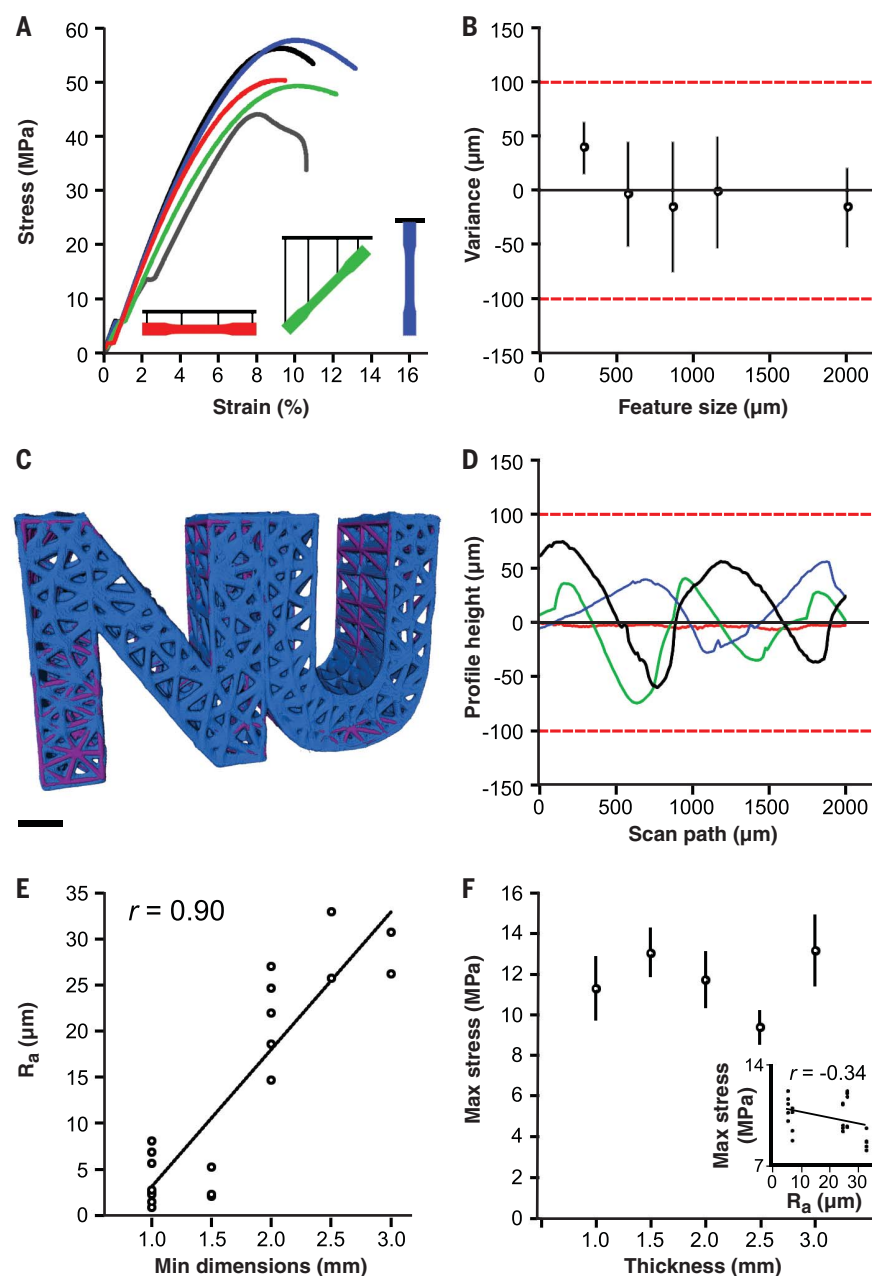


Fig. 3. Manufacturing-ready materials and resolution. (A) Type I dog-bone structures made from an ABS-like polyurethane acrylate resin exhibit isotropic mechanical properties (ASTM D638) and are comparable to a part cast from the same resin (black line) as well as injection molded ABS (gray line). (B) HARP also enables high spatial resolution and print fidelity, as evidenced by the variation between designed and printed features as a function of feature size down to $\sim 300 \mu\text{m}$ (below this, the ability to resolve parts becomes inconsistent) using a light-patterning engine with an optical resolution of $100 \mu\text{m}$. Data points are mean values; error bars represent SD across 10 5-mm posts printed in differing regions of the print bed. Red dashed lines represent the bounding constraints of ± 1 pixel for the light-patterning engine. (C) A computed tomography (CT) scan between a printed part (print rate, $120 \mu\text{m/s}$; optical resolution, $100 \mu\text{m}$) and its CAD design file reveals a volumetric correlation of 93%. Blue and purple volumes represent the printed object and the CAD object, respectively. Scale bar, 1 cm. (D) Representative height profile scans along the print direction for a series of 3-mm-thick dog bones for varying widths [red, 1 mm; green, 2 mm; blue, 4 mm; black, 6 mm; see additional data in (24) for dog bones 1 to 2.5 mm thick]. Red dashed lines represent the bounding constraints of ± 1 pixel for the light-patterning engine. (E) Analysis of the profilometry data allows for the calculation of the arithmetic surface roughness, R_a , of the printed parts as a function of minimum feature dimension and reveals a strong linear correlation with a Pearson correlation coefficient of $r = 0.90$ ($n = 20$, $r_{\text{crit}} = 0.444$ for $P = 0.05$). (F) Despite this surface roughness, the maximum tensile stress remains invariant of the feature size (data points are mean values; error bars represent SD across five Type IV dog bones, ASTM D638). The inset shows a scatterplot for each dog bone, revealing a Pearson correlation coefficient between the surface roughness and maximum tensile stress of $r = -0.34$ ($n = 25$, $r_{\text{crit}} = 0.396$ for $P = 0.05$). The hard polyurethane acrylate resin was used for all experiments in (B) to (F).

for Type I dog bones (30). The results show that the HARP-printed material maintains isotropic mechanical properties comparable to a molded part (Fig. 3A) and that these properties are independent of print orientation, unlike other modes of noncontinuous printing (fig. S8) (24). The HARP parts exhibited mechanical properties comparable to those of injection-molded ABS plastic (Fig. 3A, gray line). In addition, the HARP-printed structures could be made with high print fidelity (Fig. 3, B and C), resolving features as small as $300 \mu\text{m}$ (the theoretical optical resolution is $100 \mu\text{m}$ and is a function of the light-patterning engine) and achieving a volumetric correlation of 93% on a test print.

HARP-printed parts have a surface ridging that depends on the minimal wall thickness of the object being printed; thinner part sections result in faster resin replenishment rates and consequently yield a smoother surface. Note that this surface roughness is different from the discrete lamination layers in parts produced by traditional SLA [both digital light processing (DLP) and laser-based approaches] and is independent of the vertical resolution of the slice patterns being sent to the light-patterning engines ($10 \mu\text{m}$ z-height). To further characterize this phenomenon, we printed an array of 80 Type IV dog-bone structures of various thicknesses (1 to 3 mm) and widths

(1 to 6 mm) within the necks of the bones at a vertical print rate of $120 \mu\text{m/s}$. We then characterized the surfaces via optical microscopy and profilometry to quantify the magnitude of the profiles and the arithmetic surface roughness for each geometry. The maximum roughness, from peak to valley, was determined to be within $\pm 100 \mu\text{m}$ (i.e., ± 1 optical pixel) for all samples (Fig. 3D and fig. S9) (24), with substantially smoother surfaces for parts with a minimal dimension not exceeding 1.5 mm (within $\pm 10 \mu\text{m}$). The arithmetic surface roughness strongly correlated with the minimum part dimension (i.e., width or thickness; Fig. 3E) with a Pearson correlation coefficient of $r = 0.90$

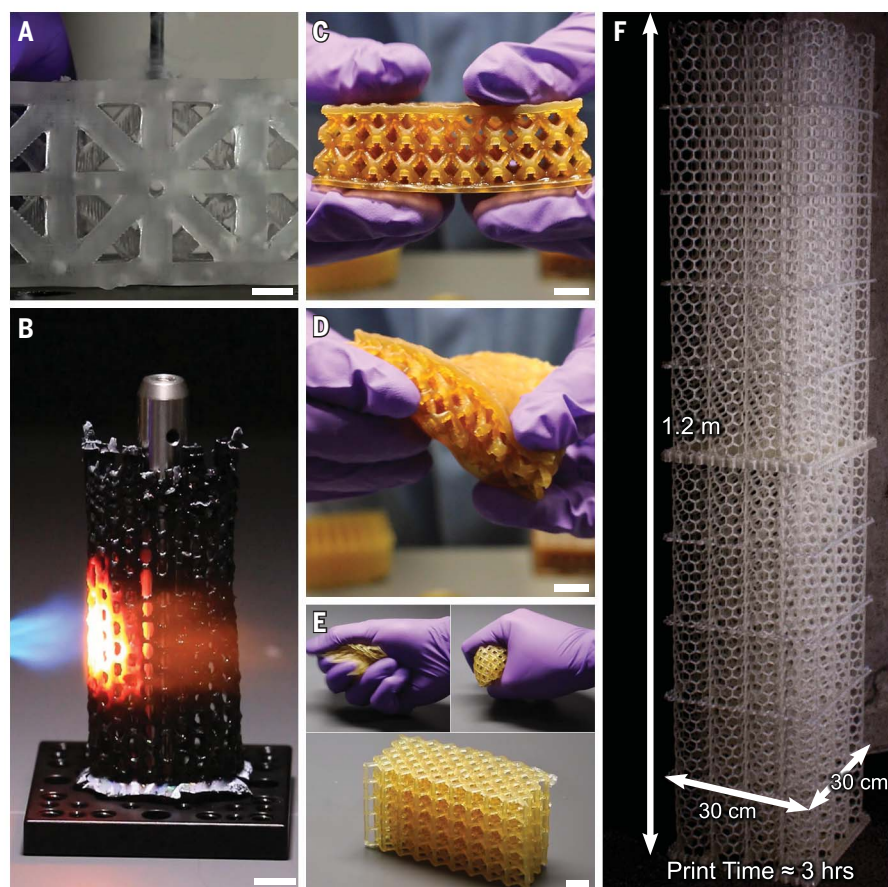


Fig. 4. A wider palette of resins. (A) A hard, machinable polyurethane acrylate part (print rate, 120 $\mu\text{m/s}$; optical resolution, 100 μm) with a hole drilled against the print direction. Traditional noncontinuous layer-by-layer printing techniques typically delaminate and fracture when drilled in this orientation. (B) A post-treated silicon carbide ceramic printed lattice (print rate of green polymer precursor, 120 $\mu\text{m/s}$; optical resolution, 100 μm) stands up to a propane torch ($\sim 2000^\circ\text{C}$). (C and D) A printed butadiene rubber structure (print rate, 30 $\mu\text{m/s}$; optical resolution, 100 μm) in a relaxed state (C) and under tension (D). (E) Polybutadiene rubber (print rate, 30 $\mu\text{m/s}$; optical resolution, 100 μm) returns to expanded lattice after compression. (F) A $\sim 1.2\text{-m}$ hard polyurethane acrylate lattice printed in less than 3 hours (vertical print rate, 120 $\mu\text{m/s}$; optical resolution, 250 μm). Scale bars, 1 cm.

($n = 20$, $r_{\text{crit}} = 0.444$ for $P = 0.05$ representing a 95% confidence level). Additionally, we printed a series of dog-bone structures (80 dog bones, five unique geometries in replicate; 120 $\mu\text{m/s}$ vertical print rate) similar to a Type IV dog bone in profile, but with various thicknesses ranging from 1 to 3 mm (a typical Type IV dog bone is 3.175 mm thick). Although differing geometries can give rise to slight variations in tensile properties, our hypothesis was that despite the varying dog-bone thickness affecting the surface roughness, the bulk material properties would be relatively invariant (Fig. 3F). Indeed, the Pearson correlation coefficient between the surface roughness and maximum tensile stress is $r = -0.34$ ($n = 25$, $r_{\text{crit}} = 0.396$ for $P = 0.05$ representing a 95% confidence level; Fig. 3F, inset), validating this hypothesis.

Taken together, these advances in throughput, mobile interface design, and materials generality solve several problems associated with large-area, large-object 3D printing. Several challenges remain, including the development of high-speed optical systems capable of retaining high lateral resolution and delivering high optical densities through the interface. Additionally, stereolithographic systems, which operate at high vertical speeds, remain limited

by the availability of low-viscosity and low-contractility resins, which can be used to generate structures with industrially relevant properties.

REFERENCES AND NOTES

- B. C. Gross, J. L. Erkal, S. Y. Lockwood, C. Chen, D. M. Spence, *Anal. Chem.* **86**, 3240–3253 (2014).
- S. H. Huang, P. Liu, A. Mokasdar, L. Hou, *Int. J. Adv. Manuf. Technol.* **67**, 1191–1203 (2013).
- J. P. Kruth, M. C. Leu, T. Nakagawa, *CIRP Ann.* **47**, 525–540 (1998).
- I. J. Petrick, T. W. Simpson, *Res. Technol. Manag.* **56**, 12–16 (2013).
- B. E. Kelly *et al.*, *Science* **363**, 1075–1079 (2019).
- M. P. de Beer *et al.*, *Sci. Adv.* **5**, eaau8723 (2019).
- R. L. Truby, J. A. Lewis, *Nature* **540**, 371–378 (2016).
- T. D. Ngo, A. Kashani, G. Imbalzano, K. T. Q. Nguyen, D. Hui, *Composites B* **143**, 172–196 (2018).
- T. Wohlers, T. Caffrey, *Manufact. Eng.* **150**, 67–73 (June 2013).
- X. Zheng *et al.*, *Science* **344**, 1373–1377 (2014).
- Y. Yagci, S. Jockusch, N. J. Turro, *Macromolecules* **43**, 6245–6260 (2010).
- P. J. Bartolo, *Stereolithography: Materials, Processes and Applications* (Springer, 2011).
- J. R. Tumbleston *et al.*, *Science* **347**, 1349–1352 (2015).
- R. Janaszewicz, J. R. Tumbleston, A. L. Quintanilla, S. J. Mecham, J. M. DeSimone, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11703–11708 (2016).
- H. Narahara, F. Tanaka, T. Kishinami, S. Igarashi, K. Saito, *Rapid Prototyp. J.* **5**, 120–128 (1999).
- C. E. Corcione, A. Greco, A. Maffezzoli, *Polym. Eng. Sci.* **46**, 493–502 (2006).
- T.-S. Wong *et al.*, *Nature* **477**, 443–447 (2011).
- S. Amini *et al.*, *Science* **357**, 668–673 (2017).
- A. Unkovskiy *et al.*, *Dent. Mater.* **34**, e324–e333 (2018).
- W. Thielicke, E. Stamhuis, *J. Open Res. Softw.* **2**, 30–39 (2014).
- P. Huang, J. S. Guasto, K. S. Breuer, *J. Fluid Mech.* **566**, 447–464 (2006).
- D. C. Tretheway, C. D. Meinhardt, *Phys. Fluids* **14**, L9–L12 (2002).
- G. Kaoullas, G. C. Georgiou, *J. Non-Newt. Fluid Mech.* **197**, 24–30 (2013).
- See supplementary materials.
- X. Kuang *et al.*, *Macromol. Rapid Commun.* **39**, e1700809 (2018).
- Z. C. Eckel *et al.*, *Science* **351**, 58–62 (2016).
- L.-A. Liew *et al.*, *Sens. Actuators A* **95**, 120–134 (2002).
- B. Husár, S. C. Ligon, H. Wutz, H. Hoffmann, R. Liska, *Prog. Org. Coat.* **77**, 1789–1798 (2014).
- S. C. Ligon, R. Liska, J. Stampfl, M. Gurr, R. Mülhaupt, *Chem. Rev.* **117**, 10212–10290 (2017).
- ASTM D638-10. Standard Test Method for Tensile Properties of Plastics (2010); www.astm.org/DATABASE.CART/HISTORICAL/D638-10.htm.

ACKNOWLEDGMENTS

We thank M. Flynn and J. Valdillez for their assistance in developing the light-patterning engines and software used to operate the printer, J. Fernandez for his contributions to understanding of the fluidic flow behavior and the CAD design of objects printed on the HARP process, and R. Davis, R. Schmidt, and E. Cottiss for their assistance in designing and analyzing resins for the printer.

Funding: This material is based on work supported by the Air Force Office of Scientific Research under award FA9550-16-1-0150 (to develop the HARP technology, to perform the velocimetry and thermal imaging, and to develop the ceramic resin); the Center for Bio-Inspired Energy Science, an Energy Frontier Research Center

funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under award DE-SC0000989 (initial UV-curable rigid urethane resin formulation); the Sherman Fairchild Foundation Inc. (initial UV-curable flexible urethane resin formulation); and a National Defense Science and Engineering Graduate Fellowship (J.L.H.). CT imaging work was performed by A. Brikha at the Northwestern University Center for Advanced Molecular Imaging, supported by NCI CCSG P30 CA060553 awarded to the Robert H. Lurie Comprehensive Cancer Center. **Author contributions:** All authors contributed to the ideas explored in

this manuscript, data analysis, and writing and editing of the manuscript. D.A.W. and J.L.H. performed experiments. C.A.M. directed the work. **Competing interests:** The authors have financial interests in Azul 3D Inc., a company that has licensed the HARP intellectual property (U.S. Patent Application 62/815,175). D.A.W., J.L.H., and C.A.M. all have secondary affiliations with Azul 3D Inc. **Data and materials availability:** All data are available in the main text or the supplementary materials. Butadiene rubber and ABS-like urethane acrylate resins require a material transfer agreement to be sampled.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/360/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Movies S1 to S8
Reference (31)

26 February 2019; accepted 20 September 2019
10.1126/science.aax1562

ORGANIC CHEMISTRY

Light-driven deracemization enabled by excited-state electron transfer

Nick Y. Shin¹, Jonathan M. Ryss², Xin Zhang¹, Scott J. Miller^{2*}, Robert R. Knowles^{1*}

Deracemization is an attractive strategy for asymmetric synthesis, but intrinsic energetic challenges have limited its development. Here, we report a deracemization method in which amine derivatives undergo spontaneous optical enrichment upon exposure to visible light in the presence of three distinct molecular catalysts. Initiated by an excited-state iridium chromophore, this reaction proceeds through a sequence of favorable electron, proton, and hydrogen-atom transfer steps that serve to break and reform a stereogenic C–H bond. The enantioselectivity in these reactions is jointly determined by two independent stereoselective steps that occur in sequence within the catalytic cycle, giving rise to a composite selectivity that is higher than that of either step individually. These reactions represent a distinct approach to creating out-of-equilibrium product distributions between substrate enantiomers using excited-state redox events.

Enantioselective reactions are essential to the pharmaceutical, agrochemical, and fine chemical industries, providing access to products enriched in just one of two mirror-image geometries. Conventional enantioselective methods either transform achiral starting materials into chiral products or rely on kinetic resolutions to differentially transform the stereoisomers of chiral reactants. Both approaches have been the subject of extensive interest and development (1). By contrast, methods for achieving selective deracemization—wherein a racemic mixture of a given compound is wholly transformed into a single enantiomer of the same molecule—are rare, despite their conceptual simplicity and potential practical benefits (Fig. 1A) (2–4). Two factors complicate the development of deracemization methods. First, the conversion of a racemic mixture into a single enantiomer is unfavorable on thermodynamic grounds because of an attendant decrease in entropy. Although this effect is small ($\Delta G^\circ = +0.42$ kcal/mol at 298 K), it requires that an additional source of energy be supplied to drive the reaction forward. The second challenge is kinetic in nature and relates to the principle of microscopic reversibility (5). As enantiomers are equal in energy by definition, any series of elementary steps along a single potential surface that converts (*S*) to (*R*) will be equally facile in the reverse direction that transforms (*R*) to (*S*). In the absence of an exogenous driving force, this necessarily results in an equilibrium (racemic) distribution of products. Accordingly, effective deracemizations require both an input of energy to impart reaction directionality and distinct mechanisms for

the elementary steps that respectively create and destroy stereochemistry.

Seminal examples from Turner, Toste, Zhou, and others have demonstrated that these requirements can be met through sequential redox transformations fueled by chemically compatible (or phase separated) oxidants and reductants, wherein the oxidation and reduction reactions occur independently and in parallel (6–8). Although effective, this approach can be challenging to generalize and requires that two stoichiometric reagents be consumed each time a molecule of substrate is processed. Excited-state reactions can also satisfy these key mechanistic requirements. Because they occur across two distinct potential energy surfaces, photochemical transformations are not subject to detailed balance and can provide access to non-Boltzmann product distributions—a benefit that underlies the success of many classical photoisomerization reactions (9, 10). Moreover, from a practical perspective, such processes require no chemical reagents, produce no stoichiometric waste, and consume nothing but photons. Bach *et al.* very recently reported groundbreaking examples of photo-driven deracemizations of allenes and cyclopropylquinolones using a chiral photosensitizer that exhibits different energy-transfer efficiencies for the two substrate enantiomers, resulting in high levels of optical enrichment (Fig. 1C) (11, 12). These and other photoisomerizations are generally understood to proceed through the electronic excited states of the substrates themselves, which strategically defines a substrate-specific paradigm (13–15).

We present here a complementary platform for light-driven deracemizations based on the use of excited-state redox events (Fig. 1B). These electron transfer-based approaches provide an alternative mechanism for driving reactions in opposition to a thermodynamic gradient yet are likely applicable to a wider

range of substrates and reaction types than direct excitation or energy transfer-based approaches. We have previously shown that excited-state redox events can be used to drive out-of-equilibrium reactions, such as intermolecular olefin hydroaminations and the isomerization of cyclic alcohols to linear ketones, wherein the reaction products are higher in energy than the starting materials (16, 17). Here, we extend these studies and describe a method for the light-driven deracemization of cyclic ureas mediated by a ternary catalyst system comprising an Ir(III)-based photoredox catalyst, a 1,1'-bi-2-naphthol (BINOL)-derived chiral phosphate base, and a cysteine-containing peptide thiol H-atom donor (Fig. 1D). This process occurs through a series of favorable electron, proton, and H-atom transfer events that serve to break and reform a stereogenic methine C–H bond. The extent of optical enrichment in these reactions is jointly determined by two independent enantioselective steps that proceed in sequence within the catalytic cycle. This results in an unusual (and beneficial) outcome in which two modestly stereoselective steps together result in an observed selectivity that is higher than that of either individual step. The discovery, optimization, scope, and a preliminary mechanistic model for this process are presented herein.

We first observed deracemization behavior serendipitously while attempting to develop an asymmetric variant of a previously reported hydroamidation reaction mediated by an Ir(III)-based photoredox catalyst, a dialkyl phosphate base, and an aryl thiol H-atom donor under visible light irradiation (Fig. 2A) (18). We found that the use of chiral BINOL phosphates as Brønsted bases was effective in this chemistry and resulted in a modest amount of enantioselectivity. However, time-course studies unexpectedly revealed that the urea product **1b** was initially formed as a racemate but became slightly optically enriched during the course of the reaction. In subsequent control reactions, we subjected racemic **1b** to the reaction conditions and observed significant optical enrichment with near-complete material recovery, indicating that a light-driven deracemization pathway was operative. Similarly, when enantiopure (*S*)-**1b** was subjected to identical conditions using an achiral phosphate base, racemization of the stereogenic C–H bond was observed.

On the basis of this discovery, we postulated that the excited state of the Ir photocatalyst reversibly oxidizes the racemic urea substrate to form a mixture of transient (and enantiomeric) arene radical cations (Fig. 2B). The stereogenic C–H bond in the resulting substrate radical cation is markedly acidified and can be deprotonated by the phosphate base to form a neutral α -amino radical (19, 20). However, as both the radical cation and the Brønsted base

¹Department of Chemistry, Princeton University, Princeton, NJ 08544, USA. ²Department of Chemistry, Yale University, New Haven, CT 06520, USA.

*Corresponding author. Email: rknowles@princeton.edu (R.R.K.); scott.miller@yale.edu (S.J.M.)

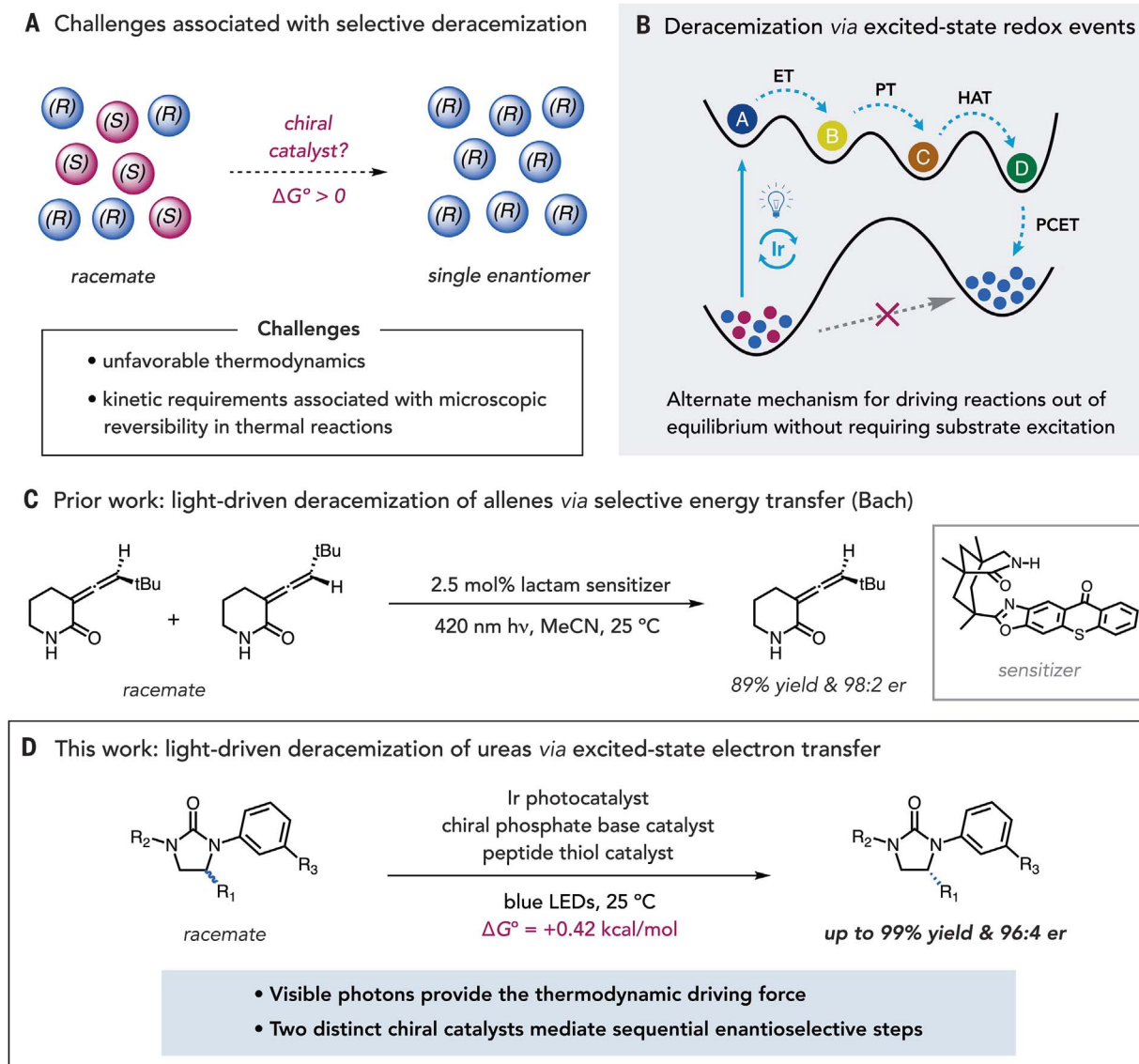


Fig. 1. Reaction development. (A) Thermodynamic and kinetic challenges in developing methods for selective deracemization. (B) General, light-driven strategies for achieving out-of-equilibrium deracemization through excited-state redox events. (C) Bach's report on light-driven deracemization through selective energy transfer. (D) Light-driven deracemization of cyclic ureas by excited-state electron transfer.

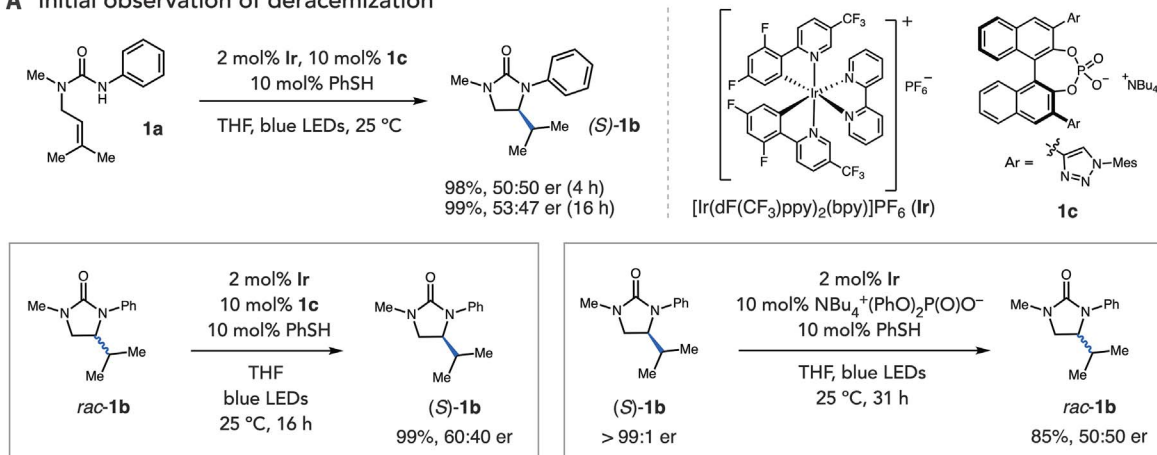
are chiral, this process serves to kinetically resolve the enantiomeric radical cations, with the fast-reacting (*R*)-enantiomer undergoing proton transfer while the slower-reacting (*S*)-enantiomer is converted back to the urea starting material by charge recombination with the reduced Ir(II) state of the photocatalyst. In this way, the reaction becomes enriched in the slower-reacting (*S*)-enantiomer. After proton transfer, the resulting α -amino radical intermediate can be reduced by H-atom transfer with the achiral aryl thiol cocatalyst to return the closed-shell urea in a nonselective process. A proton-coupled electron transfer (PCET) event among the reduced Ir(II) complex, thiol, and the protonated base could then return the active forms of all three catalysts.

If operational, this mechanism suggests that the two steps that create and destroy stereochemistry in these reactions—proton transfer and H-atom transfer—operate independently of one another and are mediated by two independent catalysts. Accordingly, when both the proton-transfer and H-atom-transfer catalysts are chiral, both elementary steps can potentially be rendered enantioselective, resulting in an unusual circumstance in which the observed stereoselectivity should be the product of the enantiomeric ratios for each of the two enantioselective steps ($er_{\text{obs}} = er_{\text{PT}} \cdot er_{\text{HAT}}$).

To evaluate this hypothesis, we elected to further study the deracemization of *N*-aryl-substituted cyclic ureas. Preliminary studies

demonstrated that a pendant amide H-bond donor group is crucial for obtaining high selectivities in the enantioselective deprotonation step using the chiral phosphate bases (vide infra), prompting us to evaluate the deracemization of urea **2a** as a model substrate (Fig. 3A). A small collection of BINOL-derived phosphate bases were explored with $[\text{Ir}(\text{dF}(\text{CF}_3)\text{ppy})_2(\text{bpy})]\text{PF}_6$ (**Ir**) and an achiral thiophenol H-atom donor catalyst in tetrahydrofuran under irradiation with blue light-emitting diodes (LEDs) at room temperature. Although a BINOL-derived phosphate with 1-adamantyl groups (**3a**) gave essentially racemic product, an analogous catalyst bearing phenyl groups gave an improved er of 69:31 (**3b**) (entries 1 and 2). We hypothesized

A Initial observation of deracemization



B Postulated mechanism for observed deracemization

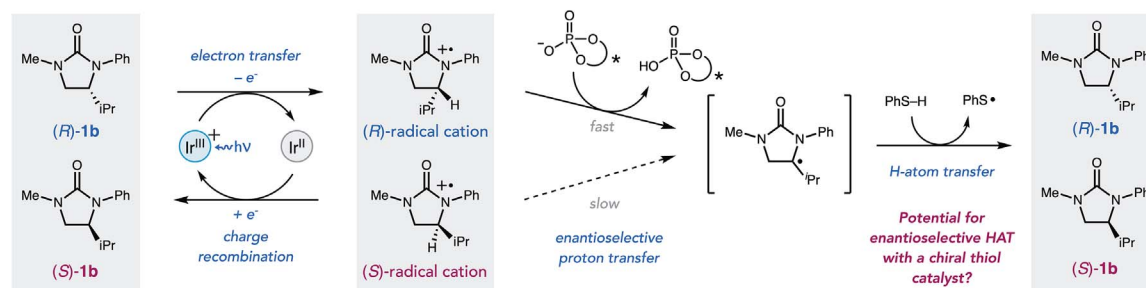


Fig. 2. Discovery of light-driven deracemization (A) Initial observations. **Ir** is racemic in all experiments. (B) Postulated mechanism.

that there might be a stabilizing π -cation interaction between the oxidized substrate radical cation and the aryl substituents of chiral phosphate, prompting us to examine catalysts bearing more expansive aryl substituents (**2I**). This in turn led to catalyst **3e**, which gave 79:21 er and good yield (entries 3 to 5). The addition of molecular sieves further improved the selectivity to 86:14 er (entry 6).

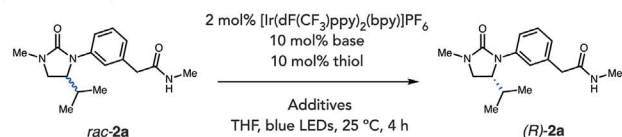
We then investigated cysteine-based oligopeptides as enantioselective H-atom-transfer catalysts. Although cysteine residues are known to mediate H-atom-transfer reactions in a variety of biological contexts, they have only occasionally been explored for use in small-molecule asymmetric catalysis, and their study for catalytic asymmetric H-atom transfer is not yet reported (22–25). A small library of tetrapeptide disulfides (which are in equilibrium with their free thiol form under the reaction conditions) were initially screened with a catalytic amount of achiral tetrabutylammonium diphenylphosphate base, providing a lead result of 68:32 er with peptide **4b** (entries 7 to 9). The corresponding thiol **4c** gave a slightly improved er of 70:30 (entry 10). Variations at the *i*+2 and *i*+3 positions demonstrated that cysteine-embedded tetrapeptide with phenylglycine as the C-terminal residue show improved er, leading to **4e** with 78:22 er and good

yield (entries 11 and 12). The use of molecular sieves slightly improved the selectivity to 79:21 er (entry 13). With optimized chiral phosphate **3e** and chiral thiol **4e** in hand, our mechanistic hypothesis predicted that reactions mediated by the stereochemically matched pair of catalysts should result in an observed er of 96:4 (86:14 • 79:21 = 96:4) (26). Reaction of *rac*-**2a** with **3e**, **4e**, **Ir**, and molecular sieves indeed produced optically enriched **2a**, with a composite selectivity of 93:7 er (entry 14) upon irradiation with blue LEDs. We speculated that the modest deviation from the predicted er value might result from a thiol-mediated racemization pathway that may become operative at high levels of optical enrichment (27). We postulated that inclusion of an alternative H-atom donor with a much weaker C–H bond might react preferentially with the alkyl thiyl radical and suppress any undesired racemization. Indeed, the addition of 50 mole % (mol %) triphenylmethane to the optimal conditions improved the er to the expected value of 96:4 (entry 15). Further investigation revealed that the use of lower catalyst loadings was also effective (entry 16). Control reactions revealed that exclusion of light or any of the reaction components resulted in either complete loss of optical enrichment or diminished levels of enantioselectivity (table S1).

With these optimized deracemization conditions in hand, we found that a variety of structural changes in the urea substrate could be accommodated (Fig. 3B). Alkyl substituents of varying steric demand could be tolerated at the stereogenic carbon with uniformly good levels of enantioselectivity (**2a** to **2e**). Substitution of the distal urea nitrogen with the free N–H amide, benzyl, or isopropyl groups also provided the desired deracemized products with high levels of stereoselectivity (**2f** to **2h**). Structural changes on the acyclic amide moiety were also tolerated, as was a benzamide derivative (**2i** to **2n**). However, an *N,N*-dimethyl amide variant demonstrated a noticeable decrease in the er (**2o**). Upon reaching the steady-state level of optical enrichment, all of the substrates studied here (**2a** to **2o**) can be recovered in nearly quantitative yield.

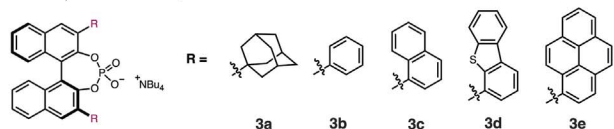
Numerous observations are consistent with the mechanistic proposal outlined above (Fig. 4A). Steady-state Stern–Volmer quenching studies and time-correlated single-photon counting experiments revealed that electron transfer between the urea substrate ($E_{p/2} = 0.91$ V versus Fc⁺/Fc in MeCN; fig. S1) and the excited state of **Ir** ($^*E_{1/2} = 0.94$ V versus Fc⁺/Fc in MeCN) (28) is kinetically rapid [the rate constants for electron transfer events (k_{ET}) for (R)-**2a** and (S)-**2a** are $9.0(8) \times 10^8$ M⁻¹s⁻¹

A Reaction optimization

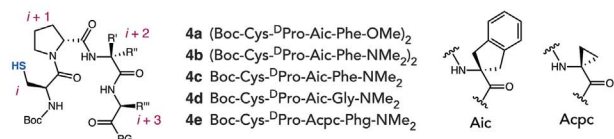


entry	base	thiol	additives	yield (%)	er
1	3a	PhSH	-	40	51:49
2	3b	PhSH	-	61	69:31
3	3c	PhSH	-	73	76:24
4	3d	PhSH	-	83	75:25
5	3e	PhSH	-	84	79:21
6*	3e	PhSH	MS	92	86:14
7†	NBu ₄ ⁺ (PhO) ₂ P(O)O ⁻	(Boc-Cys-OMe) ₂	-	92	47:53
8†	NBu ₄ ⁺ (PhO) ₂ P(O)O ⁻	4a	-	99	59:41
9†	NBu ₄ ⁺ (PhO) ₂ P(O)O ⁻	4b	-	93	68:32
10	NBu ₄ ⁺ (PhO) ₂ P(O)O ⁻	4c	-	97	70:30
11	NBu ₄ ⁺ (PhO) ₂ P(O)O ⁻	4d	-	90	72:28
12	NBu ₄ ⁺ (PhO) ₂ P(O)O ⁻	4e	-	97	78:22
13*	NBu ₄ ⁺ (PhO) ₂ P(O)O ⁻	4e	MS	95	79:21
14*	3e	4e	MS	96	93:7
15*	3e	4e	50 mol% Ph ₃ CH, MS	85	96:4
16‡	3e (5 mol%)	4e (5 mol%)	25 mol% Ph ₃ CH, MS	96	96:4

Chiral Phosphate Base Catalysts



Peptide Catalysts



B Scope studies

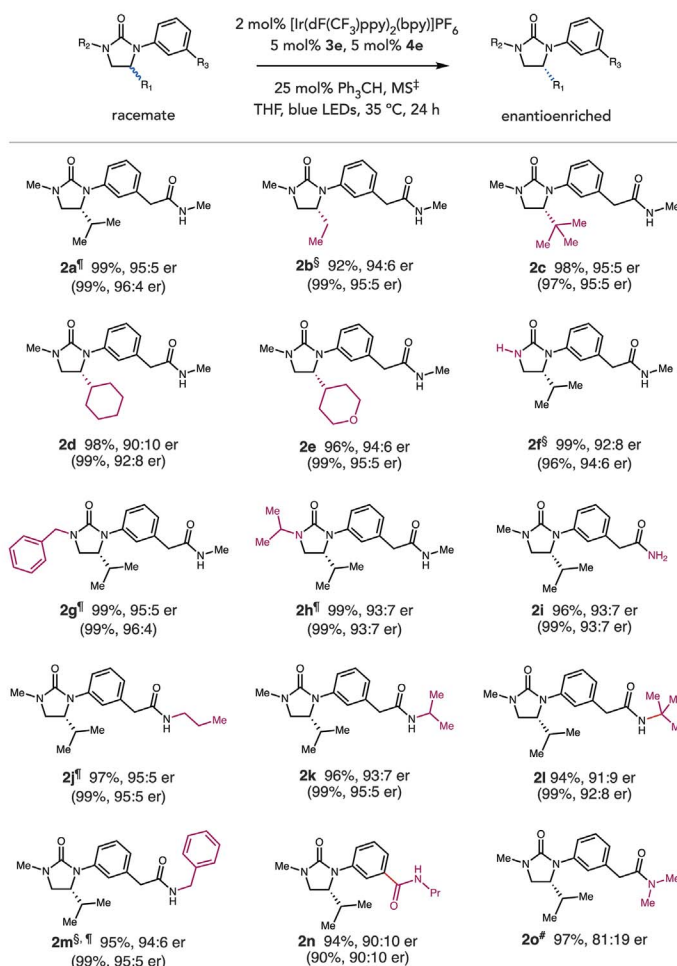


Fig. 3. Reaction optimization and scope studies. (A) Optimization of reaction conditions. Reactions were performed on a 0.025 mmol scale. Yields were determined by ¹H-nuclear magnetic resonance (NMR) analysis of crude reaction mixtures relative to an internal standard. The er was determined by high-performance liquid chromatography (HPLC) analysis on a chiral stationary phase. *10% m/v of molecular sieves (MS). †5 mol % of disulfide. ‡5% m/v of MS.

(B) Reaction scope. Reactions were run at a 0.25 mmol scale unless otherwise noted. Yields and er values are for isolated material after purification and are the average of two experiments. In parentheses are yields and er's obtained on a 0.025 mmol scale analyzed by ¹H-NMR and HPLC analysis, in which the internal reaction temperature was measured to be 25°C. §Reaction scale, 0.10 mmol. ‡Reaction time, 12 hours. #Reaction scale, 0.025 mmol, NMR yield.

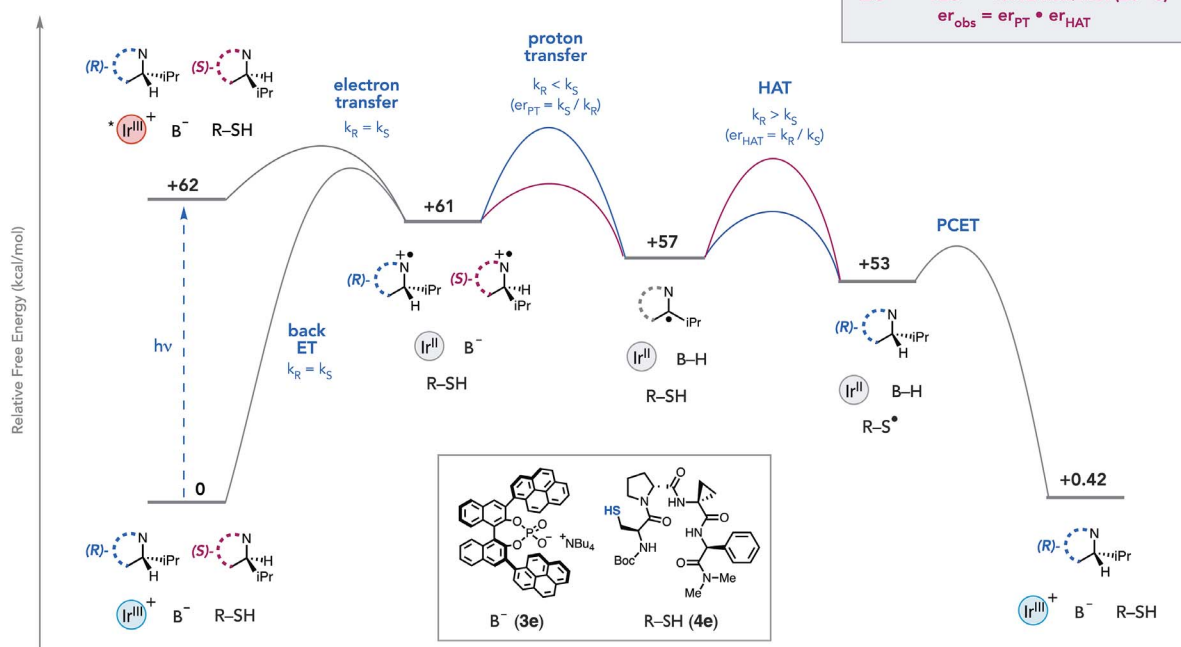
and $8.8(7) \times 10^8 \text{ M}^{-1}\text{s}^{-1}$, respectively; fig. S2]. The subsequent proton transfer, H-atom transfer, and PCET steps are also thermodynamically favorable (fig. S3). These findings reinforce the notion that, as all the elementary steps proceeding from ^{*}**Ir** are exergonic, the observed product distributions are kinetically controlled and fully decoupled from the energetic difference between the racemic starting material and the optically enriched product. The steady-state er is achieved within only 1.5 hours, and no degradation of either yield or er was observed upon extended reaction times, suggesting that the system establishes a stable nonequilibrium state (Fig. 4B). The quantum yield of this process was measured to be 4.8(3)% (fig. S5) (29). We also found that catalyst-

controlled stereoinversion from optically pure (*S*)-**2a** to (*R*)-**2a** could be achieved under the optimized condition with excellent er and reaction efficiency (Fig. 4C).

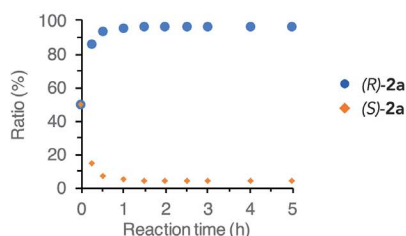
To investigate the generality of proposed synergistic stereoselectivity between the two chiral catalysts, the enantioselectivity of each catalyst was explored individually for a selected set of substrates (Table 1). First, substrates were subjected to deracemization with chiral base **3e** and thiophenol as an achiral H-atom transfer catalyst, conditions that will directly report on the er of the proton-transfer step. The observed er was generally consistent for substrates bearing a pendant amide group (entries 1 to 10). Methylation of the amide N-H or removal of the pendant amide group

decreased the er (entries 11 and 12), which supports a potential interaction between the distal amide N-H bond and the phosphate base during the asymmetric proton-transfer step. The same substrates were then deracemized with cysteine-embedded peptide **4e** and tetrabutylammonium diphenylphosphate as an achiral Brønsted base catalyst, providing a measure of the selectivity in the H-atom transfer step. Although substitution of the urea backbone led to small variations, the enantioselectivity in the HAT step is similar for all substrates with or without the pendant amide group. When the stereochemically matched forms of both chiral catalysts were used [e.g., the phosphate selectively ablates the (*S*) enantiomer and the thiol preferentially

A Energetics of light-driven deracemization



B Time-course studies



C Selective stereoinversion

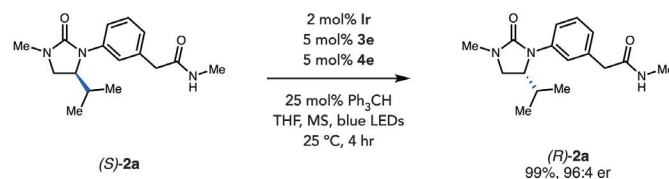


Fig. 4. Preliminary mechanistic studies. (A) Free-energy profile of light-driven deracemization from *rac*-2a to (R)-2a. Details are included in the supplementary materials (fig. S3). (B) Time-course studies for deracemization of *rac*-2a to (R)-2a. (C) Selective stereoinversion of (S)-2a to (R)-2a.

Table 1. Studies on synergistic enantioselectivity of each chiral catalyst and synergistic stereoselectivity.

Entry	Substrate	Experimental er (er_{PT}), chiral base only	Experimental er (er_{HAT}), chiral thiol only	Predicted er ($er_{\text{PT}} \cdot er_{\text{HAT}}$)	Experimental er (er_{obs}), chiral base + chiral thiol
1	2a	86:14	79:21	96:4	96:4
2	2a	86:14	21:79 (<i>ent</i> -4e)	62:38	53:47
3	2b	77:23	85:15	95:5	95:5
4	2d	80:20	76:24	93:7	92:8
5	2f	89:11	69:31	95:5	94:6
6	2g	88:12	77:23	96:4	96:4
7	2h	87:13	69:31	94:6	93:7
8	2j	85:15	81:19	96:4	95:5
9	2k	85:15	79:21	96:4	95:5
10	2m	83:17	83:17	96:4	95:5
11	2o	58:42	79:21	84:16	81:19
12	1b	49:51	77:23	76:24	74:26

Reaction conditions: 2 mol % **Ir**, 10 mol % **3e**, 10 mol % PhSH , MS, THF, blue LEDs, 25°C, 4 hours (column 3), 2 mol % **Ir**, 10 mol % $\text{NBu}_4^+(\text{PhO})_2\text{P}(\text{O})\text{O}^-$, 10 mol % **4e** (or *ent*-4e), 50 mol % Ph_3CH , MS, THF, blue LEDs, 25°C, 4 hours (column 4), 2 mol % **Ir**, 5 mol % **3e**, 5 mol % **4e** (or *ent*-4e), 25 mol % Ph_3CH , MS, THF, blue LEDs, 25°C, 4 hours (column 6). The reaction yields in all cases are >90%. Detailed experimental results are included in the supplementary materials (table S7). er_{PT} , proton-transfer er; er_{HAT} , H-atom-transfer er; er_{obs} , observed er.

reforms the (*R*) enantiomer], the resultant *er* value closely matched the predicted value in all cases. Similarly, the combination of mismatched thiol *ent-4e* with base **3e** provided diminished *er* for **2a**, indicating conflicting stereochemical preferences wherein the deprotonation and HAT events both favor the (*S*) enantiomer (entry 2). These observations are consistent with the proposed mechanism and highlight the synergistic role of the two chiral catalysts in this transformation.

We anticipate that the mechanistic features underlying this work are general and may be adapted to a wide variety of other light-driven transformations to provide nonequilibrium product distributions in a catalyst-controlled fashion.

REFERENCES AND NOTES

- E. N. Jacobsen, A. Pfaltz, H. Yamamoto, *Comprehensive Asymmetric Catalysis*, vols. I to III (Springer, 1999).
- A. J. Carnell, "Stereo-inversions using microbial redox-reactions" in *Biotransformations: Advances in Biochemical Engineering/Biotechnology*, vol. 63, K. Faber, Ed. (Springer, 1999), pp. 57–72.
- S. Servi, D. Tessaro, G. Pedrocchi-Fantoni, *Coord. Chem. Rev.* **252**, 715–726 (2008).
- W. Kroutil, K. Faber, *Tetrahedron Asymmetry* **9**, 2901–2913 (1998).
- D. G. Blackmond, *Angew. Chem. Int. Ed.* **48**, 2648–2654 (2009).
- A. D. Lackner, A. V. Samant, F. D. Toste, *J. Am. Chem. Soc.* **135**, 14090–14093 (2013).
- M. Alexeeva, A. Enright, M. J. Dawson, M. Mahmoudian, N. J. Turner, *Angew. Chem. Int. Ed.* **41**, 3177–3180 (2002).
- Y. Ji, L. Shi, M.-W. Chen, G.-S. Feng, Y.-G. Zhou, *J. Am. Chem. Soc.* **137**, 10496–10499 (2015).
- M. Kathan, S. Hecht, *Chem. Soc. Rev.* **46**, 5536–5550 (2017).
- R. D. Astumian, *Faraday Discuss.* **195**, 583–597 (2016).
- A. Hölzl-Hobmeier *et al.*, *Nature* **564**, 240–243 (2018).
- A. Tröster, A. Bauer, C. Jandl, T. Bach, *Angew. Chem. Int. Ed.* **58**, 3538–3541 (2019).
- J. B. Metternich, R. Gilmour, *J. Am. Chem. Soc.* **137**, 11254–11257 (2015).
- K. Singh, S. J. Staig, J. D. Weaver, *J. Am. Chem. Soc.* **136**, 5275–5278 (2014).
- J. B. Metternich, R. Gilmour, *J. Am. Chem. Soc.* **138**, 1040–1045 (2016).
- E. Ota, H. Wang, N. L. Frye, R. R. Knowles, *J. Am. Chem. Soc.* **141**, 1457–1462 (2019).
- A. J. Musacchio *et al.*, *Science* **355**, 727–730 (2017).
- D. C. Miller, G. J. Choi, H. S. Orbe, R. R. Knowles, *J. Am. Chem. Soc.* **137**, 13492–13495 (2015).
- J. W. Beatty, C. R. Stephenson, *Acc. Chem. Res.* **48**, 1474–1484 (2015).
- A. McNally, C. K. Prier, D. W. MacMillan, *Science* **334**, 1114–1117 (2011).
- C. R. Kennedy, S. Lin, E. N. Jacobsen, *Angew. Chem. Int. Ed.* **55**, 12596–12624 (2016).
- J. M. Ryss, A. K. Turek, S. J. Miller, *Org. Lett.* **20**, 1621–1625 (2018).
- C. E. Aroyan, S. J. Miller, *J. Am. Chem. Soc.* **129**, 256–257 (2007).
- T. Hashimoto, Y. Kawamata, K. Maruoka, *Nat. Chem.* **6**, 702–705 (2014).
- M. B. Haque, B. P. Roberts, D. A. Tocher, *J. Chem. Soc., Perkin Trans. 1* (17): 2881–2890 (1998).
- S. Masamune, W. Choy, J. S. Petersen, L. R. Sita, *Angew. Chem. Int. Ed.* **24**, 1–30 (1985).
- Y. Y. Loh *et al.*, *Science* **358**, 1182–1187 (2017).
- D. Hanss, J. C. Freys, G. Bernardinelli, O. S. Wenger, *Eur. J. Inorg. Chem.* **32**, 4850–4859 (2009).
- C. G. Hatchard, C. A. Parker, *Proc. R. Soc. London Ser. A* **235**, 518–536 (1956).

ACKNOWLEDGMENTS

We thank C. B. Roos, H. H. Ripberger, and B. C. Lainhart for preliminary studies and the synthesis of catalysts. We also acknowledge A. J. Metrano, C. R. Shugrue, and A. L. Featherston for helpful discussions, as well as Z. S. D. Toa and Q. Zhu for technical assistance. R.R.K. thanks D. MacMillan for many stimulating discussions about deracemization over the past 12 years. **Funding:** Support for this work was provided by NIH R01 GM120530 (R.R.K.) and NIH R35 GM132092 (S.J.M.). J.M.R. acknowledges the Department of Defense (DoD) for funding through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. **Author contributions:** All authors contributed to the conceptualization of the project, the planning of experiments, and interpretation of results. R.R.K. and S.J.M. directed the research. N.Y.S., J.M.R., and X.Z. conducted the experiments. N.Y.S. and R.R.K. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All experimental data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/364/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S8
References (30–46)

30 May 2019; accepted 30 August 2019
10.1126/science.aay2204

QUANTUM COMPUTING

Deterministic generation of a two-dimensional cluster state

Mikkel V. Larsen*, Xueshi Guo, Casper R. Breum, Jonas S. Neergaard-Nielsen, Ulrik L. Andersen*

Measurement-based quantum computation offers exponential computational speed-up through simple measurements on a large entangled cluster state. We propose and demonstrate a scalable scheme for the generation of photonic cluster states suitable for universal measurement-based quantum computation. We exploit temporal multiplexing of squeezed light modes, delay loops, and beam-splitter transformations to deterministically generate a cylindrical cluster state with a two-dimensional (2D) topological structure as required for universal quantum information processing. The generated state consists of more than 30,000 entangled modes arranged in a cylindrical lattice with 24 modes on the circumference, defining the input register, and a length of 1250 modes, defining the computation depth. Our demonstrated source of two-dimensional cluster states can be combined with quantum error correction to enable fault-tolerant quantum computation.

Quantum computing represents a new paradigm for information processing that harnesses the inherent nonclassical features of quantum physics to find solutions to problems that are computationally intractable on classical processors (1). In measurement-based, or cluster state, quantum computing (MBQC), the processing is performed by simple single-site measurements on a large entangled cluster state (2). This constitutes a simplification over the standard gate-based model of quantum computing, as it replaces complex coherent unitary dynamics with simple projective measurements. However, one of the outstanding challenges in realizing cluster state computation is the reliable, deterministic, and scalable generation of nonclassical entangled states suitable for universal information processing.

Several candidate platforms for scalable cluster state generation have been proposed and some experimentally realized, including solid-state superconducting qubits (3), trapped ion qubits (4, 5), and photonic qubits or qumodes, in which qubits can be encoded, generated by parametric down-conversion (6–9) or by quantum dots (10). However, none of these implementations has demonstrated true scalability combined with computational universality. The largest cluster state generated to date is a temporally multiplexed photonic state comprising entangled modes in a long chain, which, however, does not allow for universal computation because of its one-dimensional (1D) topological structure (8, 11). To achieve universality, the dimension of the cluster state must be at least two. Several proposals for generating 2D cluster states in different systems have been

proposed (12–15), but owing to technical challenges, scalable and computationally universal cluster states have yet to be produced in any physical system.

We propose and demonstrate a highly scalable scheme for the generation of cluster states for universal quantum computation based on quantum continuous variables (CVs) where information is encoded in the position or momentum quadratures of photonic harmonic oscillators (16). We use a temporally multiplexed source of optical Einstein-Podolsky-Rosen (EPR) states (17) to generate a long string of entangled modes that is curled up and fused to form a 2D cylindrical array of entangled modes. Specifically, we generate a massive cluster state of more than 30,000 entangled modes comprising an input register of $2 \times 12 = 24$ modes on which the input state may be encoded, and a length of 1250 modes for encoding operations by projective measurements, limited only by the phase stability of our setup. In addition to being universal and deterministically generated, the source is operated under ambient conditions in optical fibers at the low-loss telecom wavelength of 1550 nm. These favorable operational conditions and specifications facilitate further upscaling of the entangled state, as well as its use in applications and fundamental studies.

The canonical approach to CV cluster state generation is to apply two-mode controlled-Z gates onto pairs of individually prepared eigenstates of the momentum (or phase quadrature) operators \hat{p}_i, \hat{p}_j in adjacent modes i, j . The gate is described by the unitary operation $\hat{C}_Z = e^{ig\hat{x}_i\hat{x}_j}$, where \hat{x}_i, \hat{x}_j are the position (amplitude quadrature) operators of mode i and j , and g is the interaction strength. Applying this gate to two modes leads to entanglement in the form of quantum correlations of the two modes' quadratures. The operations and resulting state can be represented by a graph in which the nodes represent the momentum

eigenstates and the edges (links) between the nodes represent the application of a controlled-Z operation, where the interaction strength is given by the edge weight. In a practical implementation, the unphysical momentum eigenstates are replaced by highly squeezed states, while the controlled-Z operations can be imitated by phase shifts and beam-splitter transformations. To enable scalability, it has been suggested that multiplexing of spatial modes (18), frequency modes (19, 20), or temporal modes (13, 21) be used. For example, Menicucci suggested using temporal multiplexing to form a 2D cluster state combining four squeezed state generators, five beam splitters, and two delay lines (13).

We propose a simpler approach to 2D cluster state generation that lowers the experimental requirements (Fig. 1). The state is produced in four steps: (i) Pairs of squeezed vacuum states are generated at 1550-nm wavelength from two bowtie-shaped optical parametric oscillators (OPOs) by parametric down conversion (22). The states are defined in consecutive temporal modes of duration τ of the continuously generated OPO output. (ii) The squeezed vacuum pairs in spatial modes A and B are interfered on a balanced beam splitter (denoted BS_1). This produces a train of pairwise EPR-entangled temporal modes that exhibit quantum correlation between the position and momentum quadratures. Each EPR pair can be represented by a simple graph of a single edge connecting two nodes. (iii) A 1D cluster state is formed by delaying one arm of the interferometer by τ with respect to the other arm and interfering the resulting time-synchronized modes on another balanced beam splitter (denoted BS_2). The interference entangles EPR pairs along an indefinitely long chain, creating a 1D graph. (iv) In the final step, the 2D cluster state is produced by introducing another delay to one interferometer arm of duration $N\tau$ and interfering the resulting time-synchronized modes on a final beam splitter (denoted BS_3). This effectively curls up the graph and fuses the modes into an indefinitely long cylinder with N nodes on the circumference, as illustrated in Fig. 1 for $N = 12$, leading to $2 \times N = 24$ input modes distributed on the two spatial modes A and B . For a detailed description of experimental implementations, see material and methods (23).

All states and operations involved are Gaussian, meaning they can be described by Gaussian distributions of the quadrature variables in phase space. In the formalism of graphical calculus for Gaussian states (24), the generated graphs are so-called H -graphs, as they can be generated from vacuum by a single Hamiltonian and have an edge weight of $g = i \sinh(2r)G$, where r is the squeezing parameter of the two squeezing operations and $G = -1$ for the EPR states, $\pm 1/2$ for the 1D graph, and $\pm 1/4, 1/2$ for the 2D graph. Owing to the particular structure

Center for Macroscopic Quantum States (bigQ), Department of Physics, Technical University of Denmark, Fysikvej, 2800 Kgs. Lyngby, Denmark.

*Corresponding author. Email: mivila@fysik.dtu.dk (M.V.L.); ulrik.andersen@fysik.dtu.dk (U.L.A.)

of the H -graph generated here [it is self-inverse and bipartite; see supplementary text section 1.1 for details (23)], it can be transformed into a cluster state by $\pi/2$ rotations in phase space, leading to real edges of weight $g = \tanh(2r)G \rightarrow G$ for $r \rightarrow \infty$. Finally, as the $\pi/2$ phase space rotations can be absorbed into the measurement basis, or simply by appropriate redefinitions of quadratures on the rotated modes, the generated H -graph state and its corresponding cluster state are completely equivalent. [See supplementary text section 1.2 for details on the cluster state generation scheme (23).]

The produced cylindrical 2D cluster state can be shown to be a universal resource for quantum computing: In Fig. 2, the generated cylindrical cluster state is unfolded and projected into a square lattice by projective measurements in the position basis and $\pi/2$ phase-space rotations of different modes. Such

a square lattice is a well-known universal resource for quantum computing (25), and thus the initial cylindrical cluster state is itself universal. For computation, it is not necessary to project the generated cluster state into a square lattice—rather, one would in general optimize the detector settings required for the gate to be implemented. For instance, with proper settings, the cluster state can be projected into 1D dual-rail wires along the cylinder, an efficient resource for one-mode computation (8, 21) and with possible two-mode interactions between them [for details, see supplementary text section 1.4 (23)]. Doing so requires fast control of the measurement bases in between temporal modes, whereas in this work, the cluster state is measured in fixed bases for state verification.

Multipartite cluster state inseparability can be witnessed through the measurement of the uncertainties of the state nullifiers—linear com-

binations of position and momentum operators for which the cluster states are eigenstates with eigenvalue 0. For example, for the ideal two-mode EPR state, the well-known nullifiers are $\hat{n}_{\text{EPR}}^x = \hat{x}_A - \hat{x}_B$ and $\hat{n}_{\text{EPR}}^p = \hat{p}_A + \hat{p}_B$ because $\hat{n}_{\text{EPR}}^x|\text{EPR}\rangle = 0$ and $\hat{n}_{\text{EPR}}^p|\text{EPR}\rangle = 0$. For our 2D cluster state, $|2D\rangle$, the nullifiers consist of eight modes and are given by

$$\hat{n}_k^x = \hat{x}_k^A + \hat{x}_k^B - \hat{x}_{k+1}^A - \hat{x}_{k+1}^B - \hat{x}_{k+N}^A - \hat{x}_{k+N}^B + \hat{x}_{k+N+1}^A + \hat{x}_{k+N+1}^B \quad (1)$$

$$\hat{n}_k^p = \hat{p}_k^A + \hat{p}_k^B + \hat{p}_{k+1}^A + \hat{p}_{k+1}^B - \hat{p}_{k+N}^A - \hat{p}_{k+N}^B + \hat{p}_{k+N+1}^A + \hat{p}_{k+N+1}^B \quad (2)$$

as $\hat{n}_k^x|2D\rangle = 0$ and $\hat{n}_k^p|2D\rangle = 0$ [derived in supplementary text section 1.3 (23)], where the subscript indicates the temporal mode

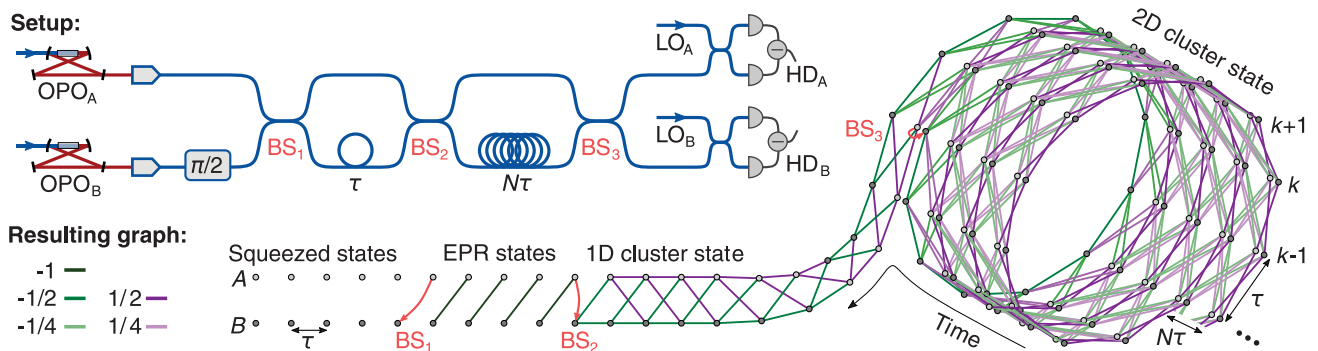


Fig. 1. Scheme of 2D cluster state generation. Squeezing is produced by two OPOs (OPO_A and OPO_B), and coupled into fiber with 97% coupling efficiency. There, temporal modes are interfered with fiber-coupled beam splitters to generate a 2D cluster state. The corresponding graph is shown: Temporal modes of squeezing with mode index k in two spatial modes A and B (bright and dark nodes) are interfered to generate EPR states at BS₁. The EPR pairs are entangled to form a 1D cluster state using a τ delay in mode B and BS₂, and the 1D cluster state is curled up to a 2D cluster state by another delay of $N\tau$ and BS₃. Using

homodyne detectors (HD_A and HD_B), the temporal mode quadratures are measured, from which the nullifiers are calculated. In the experimental implementation, the short delay is a 50.5-m fiber leading to temporal modes of 247-ns duration, whereas the long delay is a 606-m fiber such that $N = 12$, as shown in the illustrated graph. The temporal modes are defined by an asymmetric-shaped temporal mode function within the 247-ns duration, which filters out low-frequency noise and leads to less than 10^{-3} mode overlap (11). For more information, see material and methods (23).

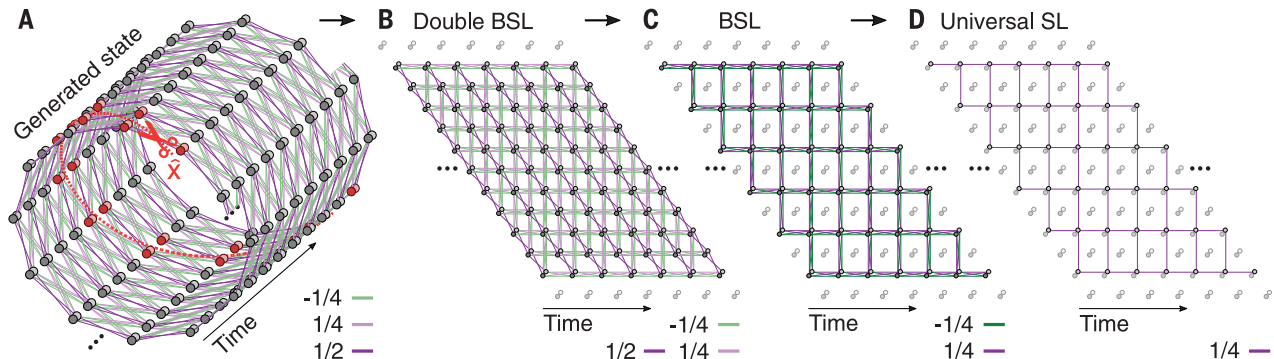


Fig. 2. Universality of the generated 2D cluster state. (A) Graph of the generated 2D cluster state. Measuring the nodes marked by red in the position basis removes all edges connected to the measured nodes, and the cylindrical graph unfolds to a plane. (B) Resulting plane 2D cluster state after the projective measurements in (A), consisting of two bilayer square lattices (double BSL)

connected by edges of weight $1/2$. (C) Single BSL after projective measurement of half the modes in (B) in the position basis. (D) Square lattice (SL) after projective position measurements of all modes in spatial mode B (dark nodes), and applying the Fourier gate ($\pi/2$ phase delay) on half the modes in spatial mode A (bright nodes). This SL is a traditional universal resource state for MBQC.

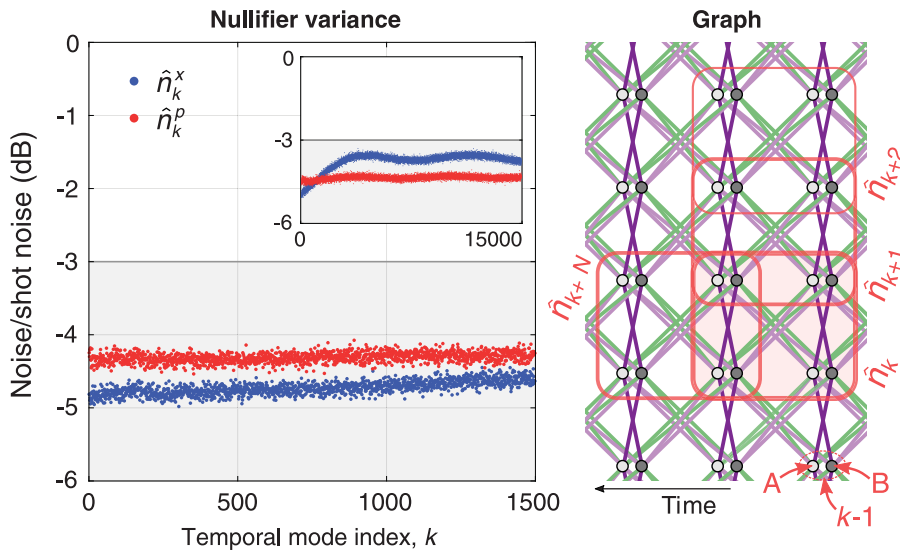


Fig. 3. Experimental result. On the right graph, the nullifiers in Eq. 1 and 2 are shown on the 2D cluster state lattice with the measured variance of 1500 consecutive nullifiers shown in the left plot. Here, the variance is calculated from 10,000 measurements of each nullifier. All nullifier variances are seen to be well below the -3 -dB inseparability bound derived in supplementary text section 2 (23), and thus the generated cluster state is completely inseparable. In the inset, the nullifier variance of a larger data set with $2 \times 15,000 = 30,000$ modes is shown. Again, with all modes below the -3 -dB inseparability bound, we conclude the successful generation of a 30,000-mode 2D cluster state. The rapid increase of the variance in \hat{n}_k^x and its periodic variation is caused by phase fluctuation of the squeezing sources, as described in supplementary text section 4 (23).

index and N is the number of temporal modes in the cluster state circumference.

The practically realizable cluster state is never an exact eigenstate of the nullifiers, because such a state is unphysical. The measurement outcomes of the nullifiers are therefore not exactly zero in every measurement but possess some uncertainties around zero. A condition for complete inseparability of the 2D cluster state [derived in supplementary text section 2 (23)] leads to a bound on the variances of all nullifiers of 3 dB squeezing below the shot noise level. Therefore, to witness full inseparability, we must observe more than 3-dB squeezing for all nullifiers. In Fig. 3, the measured nullifier variances are shown for a dataset of 1500 nullifiers, and they are all observed to be well below the -3 -dB bound; we measure an averaged variance of -4.7 and -4.3 dB for \hat{n}_k^x and \hat{n}_k^p , respectively. In the inset of Fig. 3, we present the measurement of a longer cluster state of 15,000 temporal modes, which corresponds to a measurement time of 4 ms. Although phase instabilities are clearly seen to affect the performance in terms of variations of the nullifier variances, all variances stay below the -3 -dB bound. The 2D cluster of $2 \times 15,000 = 30,000$ modes is thus fully inseparable. Not all 30,000 modes of the cluster state need to exist simultaneously when performing projective measurements for computation. Indeed, only a single temporal mode of the cluster state needs to exist while the remaining modes of

the state are under construction. Hence, the cluster state can be immediately consumed for computation while being generated, with no additional state storage necessary (see supplementary text section 1.4 for a possible measurement scheme for computation on the cluster state).

With the deterministic generation of a universal 2D cluster state, we have [in parallel with Asavanant *et al.* (26)] constructed a platform for universal MBQC. Its scalability was demonstrated by entangling 30,000 optical modes in a 2D lattice that includes 24 input modes and allows for a computation depth of 1250 modes. Because only a few modes exist simultaneously, we are not limited by the coherence time of the light source, and thus the number of operations depends only on the phase stability of the system. Unlimited computational depth can therefore be achieved by implementing continuous feedback control of the system for phase stabilization, as demonstrated for the 1D photonic cluster state in (11). The results presented here and in (26) are similar: Both 2D cluster states are generated deterministically in the CV regime with comparable size and amount of squeezing in the nullifier variance. However, with only two squeezing sources, three interference points, and operation in fiber, the experimental setup demonstrated here is simpler, whereas in (26), larger-bandwidth OPOs are demonstrated that result in shorter delay lines. In both systems,

the number of input modes can be readily increased by using OPOs with larger bandwidths, possibly combined with a longer time delay of the second interferometer. For example, using OPOs with a 1-GHz bandwidth (65 times wider) and an interferometer delay that is twice as long, a state with ~ 1500 input modes can be generated. Large-bandwidth OPOs have been demonstrated, but phase stability and losses in the delay lines are more challenging. Although phase fluctuation depends only on experimental control, which we expect to improve with continuous phase stabilization, delay losses are unavoidable, and increasing the OPO bandwidth may be a better solution than increasing the delay lengths.

CV cluster states are described by Gaussian statistics, but it is known that an element (state, operation, or measurement) of non-Gaussian quadrature statistics is required for universal quantum computing (27). Such an element could be a photon number-resolving detector (PNRD) or an ancillary cubic-phase state (21, 28). Despite recent experimental efforts in developing high-efficiency PNRD (29) and deterministically generating optical states with non-Gaussian statistics (30), the formation of the required non-Gaussianity of the cluster state still constitutes an important challenge to be tackled in the future. Another currently limiting factor to achieving quantum computation is the existence of finite squeezing in the cluster, leading to excess quantum noise and thus computational errors. However, these errors can be circumvented using Gottesman-Kitaev-Preskill (GKP) state encoding (28) concatenated with traditional qubit error correction schemes, leading to fault-tolerant computation with a 15- to 17-dB squeezing threshold (31). Another recently discovered advantage of the GKP encoding is that in addition to fault tolerance, it also allows for universality without adding extra non-Gaussian states or operations (32). Although GKP states have recently been produced in the microwave regime (33) and in trapped-ion mechanical oscillators (34), their production in the optical regime remains a task for future work. [For further discussion on quantum computation using the generated cluster state, see supplementary text section 1.4 (23).] Although a path toward fault-tolerant universal quantum computing using CV cluster states has been established, it is highly likely that the first demonstrations of CV quantum computation will be nonuniversal algorithmic subroutines such as boson sampling and instantaneous quantum computing (35). With the large, but noisy, cluster state demonstrated here, interesting future work will be to implement basic Gaussian circuits and investigate, for example, the attainable circuit depth. Furthermore, the technique of folding a 1D cluster state into a 2D structure could be extended using an additional interferometer to form 3D cluster states,

which might be suitable for topologically protected MBQC.

REFERENCES AND NOTES

1. T. D. Ladd *et al.*, *Nature* **464**, 45–53 (2010).
2. R. Raussendorf, H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
3. Y. Wang, Y. Li, Z. Yin, B. Zeng, *npj Quantum Information* **4**, 46 (2018).
4. O. Mandel *et al.*, *Nature* **425**, 937–940 (2003).
5. B. P. Lanyon *et al.*, *Phys. Rev. Lett.* **111**, 210501 (2013).
6. P. Walther *et al.*, *Nature* **434**, 169–176 (2005).
7. Y. Tokunaga, S. Kuwashi, T. Yamamoto, M. Koashi, N. Imoto, *Phys. Rev. Lett.* **100**, 210501 (2008).
8. S. Yokoyama *et al.*, *Nat. Photonics* **7**, 982–986 (2013).
9. M. Chen, N. C. Menicucci, O. Pfister, *Phys. Rev. Lett.* **112**, 120505 (2014).
10. I. Schwartz *et al.*, *Science* **354**, 434–437 (2016).
11. J. Yoshikawa *et al.*, *APL Photonics* **1**, 060801 (2016).
12. S. E. Economou, N. Lindner, T. Rudolph, *Phys. Rev. Lett.* **105**, 093601 (2010).
13. N. C. Menicucci, *Phys. Rev. A* **83**, 062314 (2011).
14. P. Wang, M. Chen, N. C. Menicucci, O. Pfister, *Phys. Rev. A* **90**, 032325 (2014).
15. R. N. Alexander *et al.*, *Phys. Rev. A* **94**, 032327 (2016).
16. C. Weedbrook *et al.*, *Rev. Mod. Phys.* **84**, 621–669 (2012).
17. M. D. Reid *et al.*, *Rev. Mod. Phys.* **81**, 1727–1751 (2009).
18. S. Armstrong *et al.*, *Nat. Commun.* **3**, 1026 (2012).
19. N. C. Menicucci, S. T. Flammia, O. Pfister, *Phys. Rev. Lett.* **101**, 130501 (2008).
20. Y. Cai *et al.*, *Nat. Commun.* **8**, 15645 (2017).
21. R. N. Alexander, S. Yokoyama, A. Furusawa, N. C. Menicucci, *Phys. Rev. A* **97**, 032302 (2018).
22. U. L. Andersen, T. Gehring, C. Marquardt, G. Leuchs, *Phys. Scr.* **91**, 053001 (2016).
23. See the online supplementary materials.
24. N. C. Menicucci, S. T. Flammia, P. van Loock, *Phys. Rev. A* **83**, 042335 (2011).
25. M. Gu, C. Weedbrook, N. C. Menicucci, T. C. Ralph, P. van Loock, *Phys. Rev. A* **79**, 062318 (2009).
26. W. Asavanant *et al.*, *Science* **366**, 373–376 (2019).
27. S. D. Bartlett, B. C. Sanders, S. L. Braunstein, K. Nemoto, *Phys. Rev. Lett.* **88**, 097904 (2002).
28. D. Gottesman, A. Kitaev, J. Preskill, *Phys. Rev. A* **64**, 012310 (2001).
29. G. Thekkadath *et al.*, arXiv:1908.04765 [quant-ph] (2019).
30. B. Hacker *et al.*, *Nat. Photonics* **13**, 110–115 (2019).
31. B. W. Walshe, L. J. Mensen, B. Q. Baragiola, N. C. Menicucci, *Phys. Rev. A* **100**, 010301 (2019).
32. B. Q. Baragiola, G. Pantaleoni, R. N. Alexander, A. Karanjai, N. C. Menicucci, arXiv:1903.00012 [quant-ph] (2019).
33. P. Campagne-Ibarcq *et al.*, arXiv:1907.12487 [quant-ph] (2019).
34. C. Flühmann *et al.*, *Nature* **566**, 513–517 (2019).
35. D. Su *et al.*, *Phys. Rev. A* **98**, 032316 (2018).
36. M. V. Larsen, X. Guo, C. R. Breum, J. S. Neergaard-Nielsen, U. L. Andersen, Deterministic generation of a two-dimensional cluster state. figshare (2019).

ACKNOWLEDGMENTS

We thank R. N. Alexander for useful discussion on the final manuscript and J. B. Brask for proofreading. **Funding:** The work was supported by the Danish National Research Foundation through the Center for Macroscopic Quantum States (bigQ, DNRF142) and the VILLUM FOUNDATION Young Investigator Programme (grant no. 10119). **Author contributions:** M.V.L. and U.L.A. conceived the project. J.S.N.-N., X.G., and C.R.B. designed and built the squeezing sources. M.V.L. developed the theoretical background, designed the experiment, and built the setup. M.V.L. performed the experiments and data analysis. The project was supervised by U.L.A. and J.S.N.-N. The manuscript was written by U.L.A., M.V.L., and J.S.N.-N. with feedback from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Experimental data and analysis code is available at figshare (36).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/369/suppl/DC1
Materials and Methods
Figs. S1 to S14
Table S1
References (37–41)

19 June 2019; accepted 25 September 2019
10.1126/science.aay4354

QUANTUM COMPUTING

Generation of time-domain-multiplexed two-dimensional cluster state

Warit Asavanant¹, Yu Shiozawa¹, Shota Yokoyama², Baramée Charoensombutamon¹, Hiroki Emura¹, Rafael N. Alexander³, Shuntaro Takeda^{1,4}, Jun-ichi Yoshikawa¹, Nicolas C. Menicucci⁵, Hidehiro Yonezawa², Akira Furusawa^{1*}

Entanglement is the key resource for measurement-based quantum computing. It is stored in quantum states known as cluster states, which are prepared offline and enable quantum computing by means of purely local measurements. Universal quantum computing requires cluster states that are both large and possess (at least) a two-dimensional topology. Continuous-variable cluster states—based on bosonic modes rather than qubits—have previously been generated on a scale exceeding one million modes, but only in one dimension. Here, we report generation of a large-scale two-dimensional continuous-variable cluster state. Its structure consists of a 5- by 1240-site square lattice that was tailored to our highly scalable time-multiplexed experimental platform. It is compatible with Bosonic error-correcting codes that, with higher squeezing, enable fault-tolerant quantum computation.

Quantum computers promise applications beyond what is possible with their classical counterparts (1). Recent work has produced advances in quantum computing with stationary matter qubits such as superconducting systems (2), ion trap systems (3), and silicon-based systems (4). With access to both high-quality qubits and high-fidelity quantum gates, it is thus believed that a small-scale quantum computer is within reach. In order to bring these platforms to the scale at which they become useful, efforts [for example, (5)] have been made to address the difficulties of preparing, interfacing, addressing, and tuning large numbers of qubits. However, a technological leap will be necessary to achieve a scalable quantum computer along these lines.

In the light of this, we sought a more direct path to scalability. Rather than sequentially preparing and interfacing qubits, one can prepare a large-scale quantum resource state whose constituent quantum systems (such as qubits or modes) are interconnected in advance. Provided that this resource—known as a cluster state (6)—has the required structure and scale, then quantum computing can proceed by means of purely local (single-site) quantum measurements. This paradigm is known as measurement-based quantum computation (MBQC) (7, 8). The scale of the cluster state

typically determines the number of possible computational steps. Its structure—codified as a graph that describes the entanglement properties of the state—determines which logic gates can be implemented with local measurements. For example, a cluster state corresponding to multiple disjoint one-dimensional (1D) graphs (Fig. 1A) can be used to implement multistep computations, with teleportation along the length of the cluster state being analogous to time evolution along a single circuit wire. However, the absence of connections between wires makes this state incapable of generating entanglement between inputs on separate wires, and hence, this resource is insufficient for universal quantum computing. A 2D square lattice graph (Fig. 1B) possesses connectivity in two directions, allowing for time evolution through teleportation in one direction and the multi-input unitary evolution (that generates entanglement) between inputs through connections in the other direction. It enables universal quantum computing by means of local measurements and hence is a universal resource state for MBQC.

Finding a universal resource state that is experimentally feasible (so that it can be generated on a large scale) is therefore of paramount importance to the development of a universal measurement-based quantum computer. Cluster states based on bosonic modes—known as continuous-variable cluster states (8)—can be generated deterministically by using compact quantum optics experiments (9). Despite several theoretical proposals for generating universal continuous-variable cluster states [for example, (9–12)], until now the current state-of-the-art experimental demonstrations have been limited to either small-scale (few-system) cluster states (13) or to the generation of large-scale 1D cluster states, which are insufficient for multi-input MBQC (14–16). Generation of even moderately sized

2D cluster states by using discrete variables (qubits) has also never been reported.

In this work, we report the generation of a large-scale universal continuous-variable cluster state. The scalability of our experimental method stems from using a time-domain multiplexing (TDM) architecture for continuous-variable (CV) optical systems (9). Optical CV quantum information is encoded within continuous-valued quadratures \hat{x}_k and \hat{p}_k that satisfy $[\hat{x}_j, \hat{p}_k] = i\hbar\delta_{jk}$ and correspond to the complex electric field amplitudes of optical modes residing in temporally localized wave packets. An unlimited number of modes can be continuously and deterministically prepared from a few sources and then transformed into a continuous-variable cluster state through the repeated use of a circuit made of only a small number of optical components. Entanglement in the resulting state can be observed from correlations in quadrature values of different temporal modes. In this way, we generate and verify a universal cluster state with a 5- by 1240-site 2D square lattice structure, with the possibility of further extension along both lattice dimensions by several orders of magnitude in the near future by use of current technology.

Our method generates a continuous-variable cluster state from four squeezed light sources and a linear optical network consisting of five beam splitters and two delay lines (Fig. 1C). After single-mode squeezed states leave each optical parametric oscillator (OPO), they are converted to two-mode squeezed states by beam splitter 1 and beam splitter 2 and then a four-mode entangled state by beam splitter 3. Next, the delay lines multiplex these in time so that they form a 2D (but still disconnected) grid layout with cylindrical boundary conditions. Last, beam splitter 4 and beam splitter 5 produce additional connections, resulting in a continuous cylindrical structure (Fig. 1D) [as shown for the case in which $N = 30$]. A more detailed description of this generation procedure can be found in (17), section SM2. The surface of this cylinder consists of sites (macronodes) that are arranged in a 2D square lattice whose temporal modes (micronodes) are connected in a nontrivial way (Fig. 1E). The delay between temporal modes arriving at each detector is set by the shorter delay Δt . The longer delay line $N\Delta t$ controls the circumference length (N sites) of the cylinder. Therefore, our method can generate arbitrarily long 2D cluster states, and only the length of the long delay line needs to be increased to extend the circumferential dimension. One limitation on N is that $N\Delta t$ must be below the coherence time of the light source.

Relative to the 1D case (14, 15), generating a 2D cluster state requires a more complex network of beam splitters. This is further complicated by stabilization issues and additional optical

¹Department of Applied Physics, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. ²Centre for Quantum Computation and Communication Technology, School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2600, Australia. ³Center for Quantum Information and Control, University of New Mexico, MSC07-4220, Albuquerque, NM 87131-0001, USA. ⁴Institute of Engineering Innovation, School of Engineering, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan. ⁵Centre for Quantum Computation and Communication Technology, School of Science, RMIT University, Melbourne, Victoria 3001, Australia. *Corresponding author. Email: akiraf@ap.t.u-tokyo.ac.jp

losses that arise from requiring the longer delay line $N\Delta t$ and adjustment issues that arise from the constraint that the length ratio of the two optical delay lines must be an integer. The former can be addressed by instead shortening the length of each temporal-mode wave packet, which allows us to make the cluster state using delay lines with shorter lengths. We achieved this by developing broadband squeezed light sources, broadband homodyne detectors, and control systems (18). With this, we reduced the width of the wave packet from ~ 160 ns in (14, 15) to $\Delta t = 40$ ns. Regarding the latter issue, the current setup was designed

so that the lengths of two optical delay lines can be easily measured and adjusted (17) while keeping the number of the optical components the same or less than that in the previous proposals (9, 10). We picked $N = 5$ for experimental demonstration.

To implement universal CV quantum computation, we require the ability to perform arbitrary multimode Gaussian operations and at least one non-Gaussian operation (19). Although it is possible to map the structure of our 2D square lattice cluster state to the more standard square lattice [whose methodology for implementing MBQC is known (8)], such a

mapping introduces excessive noise (20). We have noticed that a similar experiment of generating a 2D cluster state has recently been reported (21), although the methodology to avoid such a mapping has not been developed for their cluster state yet. Here, we describe a more efficient method to use our cluster state without such mapping. The structure of multimode Gaussian quantum circuits that can be implemented with our resource state by means of local homodyne measurements is shown in Fig. 2A. The number of input modes N is equal to the circumference length, and each gate is implemented by means of teleportation

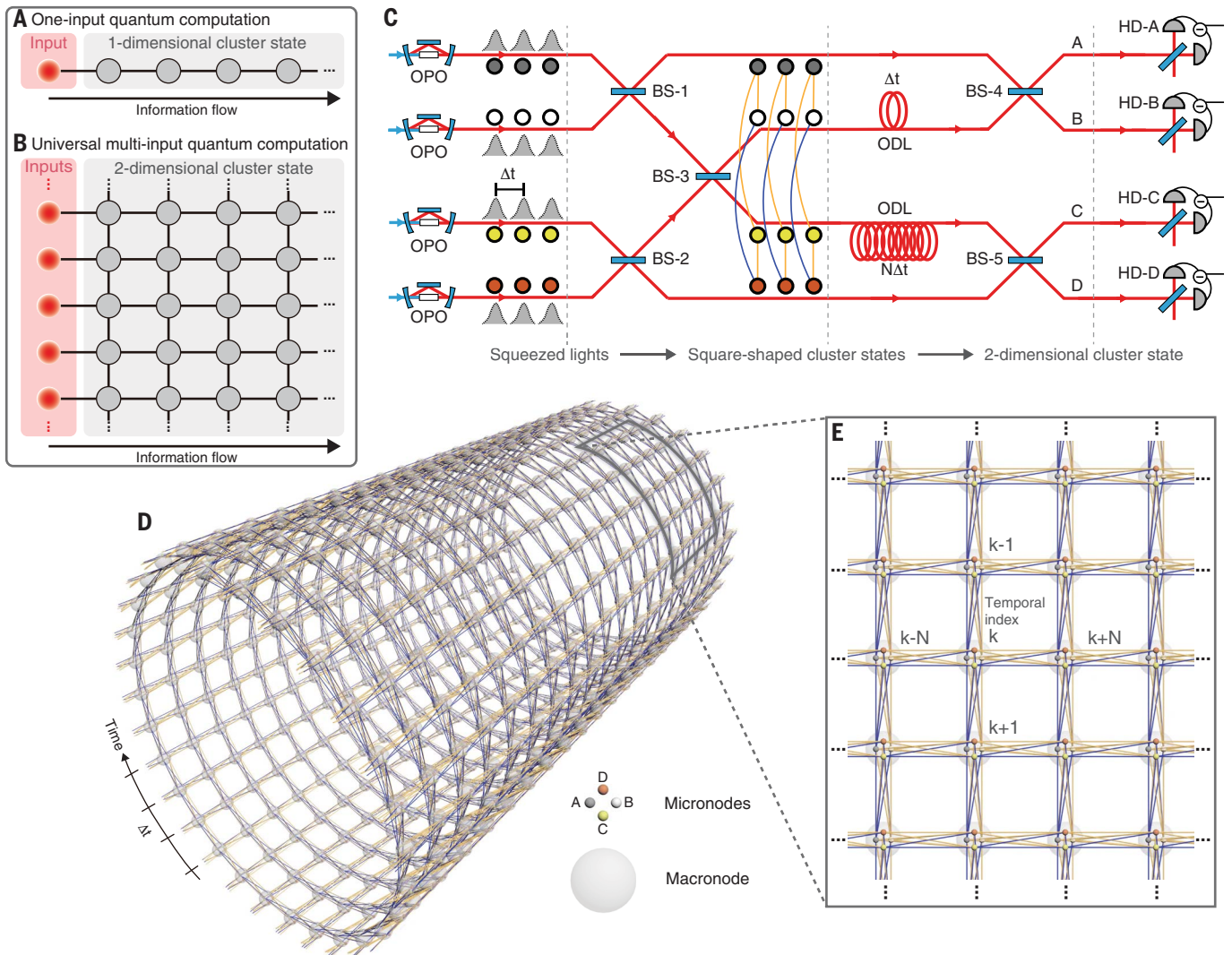


Fig. 1. MBQC and 2D cluster state. (A and B) Abstract illustration of MBQC. (A) One-input MBQC by using 1D cluster state. (B) Universal multi-input MBQC using 2D cluster state. Each colored circle represents a mode, and each link represents quantum entanglement. (C) Schematic of our experimental setup for the 2D cluster state. OPO, optical parametric oscillator; BS, beam splitter; ODL, optical delay line; Δt , time interval between adjacent wave packets; N , integer corresponding to number of inputs that 2D cluster states can take in computation; HD, homodyne detector. All beam splitters are 50:50. (D and E) 2D cluster state. (D) Example for the case in which $N = 30$.

(E) Zoom in of the state. The representations of states make use of the simplified graphical calculus (9). Each small node (small colored sphere) of the graph, which we call a micromode, represents a localized wave packet at each temporal index. Four micromodes at each temporal index k can be grouped into a single site (large gray sphere), called a macromode. The links and their colors represent how micromodes are entangled. The 2D cluster state has a helical graph structure, with N macromodes on every single turn of the helix. For actual experimental demonstration, we use $N = 5$. Full descriptions are given in (17).

along the length of the cylindrical cluster state. Further details for how to implement multimode Gaussian operations are given in (17), section SM6. Non-Gaussian operations can be implemented by replacing homodyne-C or homodyne-D with cubic-phase ancilla-assisted measurement (10). When implementing an encoded qubit-level computation by means of the Gottesman-Kitaev-Preskill (GKP) error correction scheme for CV cluster states (10, 22–24), the only non-Gaussian resource required for both universal and fault-tolerant quantum computing is GKP logical $|0\rangle$ states, which can be inserted into the cluster state at regular intervals (25), and no measurements other than homodyne measurement are required.

An n -mode Gaussian pure state $|\psi\rangle$ can be efficiently characterized by a list of n -independent linear nullifiers, which are linear combinations of the quadrature operators that have $|\psi\rangle$ as their mutual zero-eigenspace. Nullifiers also play an important role in verifying genuine

multipartite inseparability for experimentally generated states that can be generated from two-mode squeezed states by means of a sequence of beam splitters (10). Such states are approximately nullified by linear combinations of quadratures that are either all of position- or momentum-type. By measuring these operators, if the states are sufficiently highly squeezed, then genuine multipartite inseparability can be verified with the van Loock–Furusawa criterion.

Our state can be characterized in this way by measuring

$$\hat{\delta}_k^{(x,1)} = \hat{x}_k^A + \hat{x}_k^B - \frac{1}{\sqrt{2}} \left(-\hat{x}_{k+1}^A + \hat{x}_{k+1}^B + \hat{x}_{k+N}^C + \hat{x}_{k+N}^D \right) \quad (1)$$

$$\hat{\delta}_k^{(p,1)} = \hat{p}_k^A + \hat{p}_k^B + \frac{1}{\sqrt{2}} \left(-\hat{p}_{k+1}^A + \hat{p}_{k+1}^B + \hat{p}_{k+N}^C + \hat{p}_{k+N}^D \right) \quad (2)$$

$$\hat{\delta}_k^{(x,2)} = \hat{x}_k^C - \hat{x}_k^D - \frac{1}{\sqrt{2}} \left(-\hat{x}_{k+1}^A + \hat{x}_{k+1}^B - \hat{x}_{k+N}^C + \hat{x}_{k+N}^D \right) \quad (3)$$

$$\hat{\delta}_k^{(p,2)} = \hat{p}_k^C - \hat{p}_k^D + \frac{1}{\sqrt{2}} \left(-\hat{p}_{k+1}^A + \hat{p}_{k+1}^B - \hat{p}_{k+N}^C + \hat{p}_{k+N}^D \right) \quad (4)$$

where \hat{x}_k^j and \hat{p}_k^j are quadrature operators at temporal mode index k and at spatial index j . Shown in Fig. 3 are the quadrature values and quantum correlations of quadratures corresponding to each type of nullifier of the first 50 temporal mode indices. These quadrature values are obtained by processing the time-domain electrical signal from each homodyne detector and taking appropriate linear combinations. We observed strong correlations between quadrature values from different temporal modes, which are qualitative evidences of quantum entanglement of our states. By

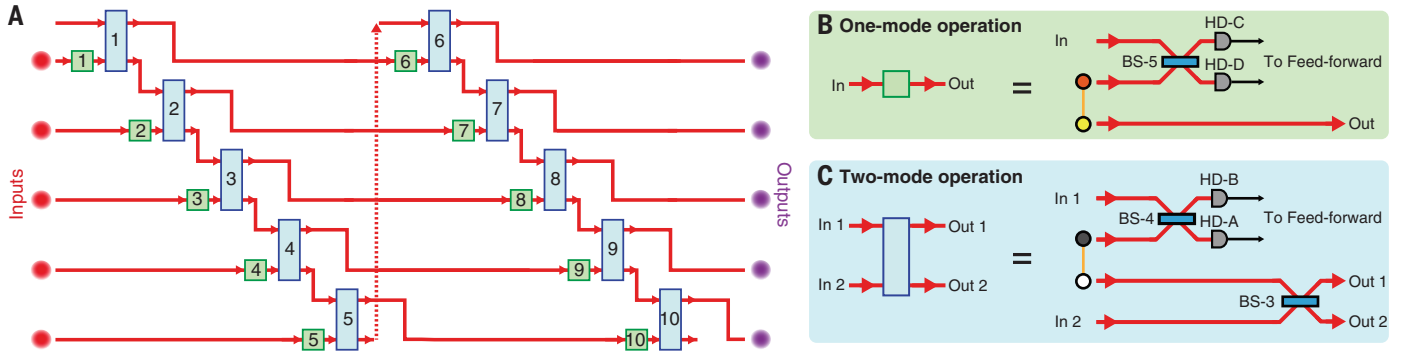


Fig. 2. Quantum computation with our cluster state. (A) Equivalent quantum circuit that is implemented when our state is used. We show the case for five inputs in which 40 light modes (10 temporal mode indices) of 2D cluster states are used. The number in each box is the index of the measured temporal modes. (B and C) The circuit is composed of multiple

quantum teleporters: (B) one-mode operation and (C) two-mode operation. Two-mode operations can be turned off by selecting the same measurement basis for HD-A and HD-B. Classical feed-forward does not have to be implemented immediately after homodyne measurements and can be delayed to the end of the computation for Gaussian-only computations.

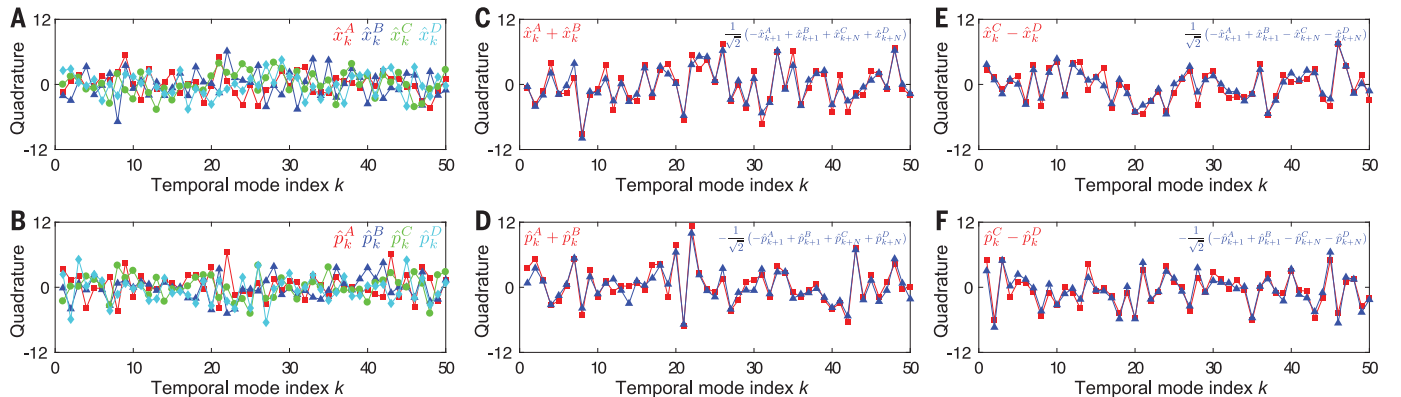


Fig. 3. Quadrature values and four types of quadrature correlations of the first 50 temporal mode indices. (A and B) Single-shot quadrature values of \hat{x}_k^j and \hat{p}_k^j obtained by processing time-domain signals from homodyne detector- j ($j = A, B, C, D$). (C to F) Correlations of quadrature values corresponding to $\hat{\delta}_k^{(x,1)}$, $\hat{\delta}_k^{(p,1)}$, $\hat{\delta}_k^{(x,2)}$, and $\hat{\delta}_k^{(p,2)}$, respectively. Although the quadrature values measured at each

homodyne detector seemed to be just fluctuating randomly around zero, we observed four types of strong quantum correlations between six quadrature values with different temporal mode index k and spatial index j . The quadrature values are plotted by using the unit at which variance of the vacuum state is equal to 1.

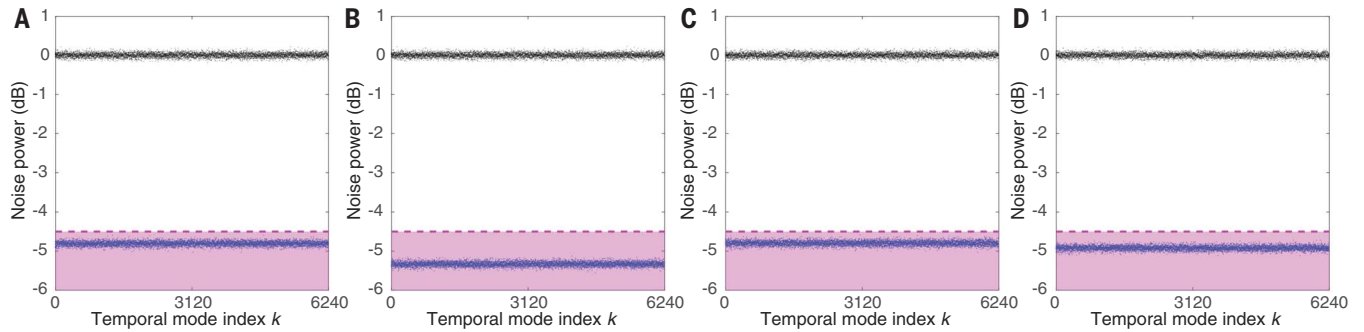


Fig. 4. Verification of generated cluster state for 24,960 temporal modes. (A to D) Measurement results for each type of nullifier: $\hat{\delta}_k^{(x,1)}$, $\hat{\delta}_k^{(p,1)}$, $\hat{\delta}_k^{(x,2)}$, and $\hat{\delta}_k^{(p,2)}$, respectively. Black points indicate measured variances of shot noise, which are used as reference levels. Blue points indicate measured variances of nullifiers. Purple regions indicate regions where the variances of the nullifiers are below -4.5 dB compared with shot noise, which indicate entanglement. The variances of four types of nullifiers satisfied quantum inseparability criteria up to $k = 6240$ corresponding to $4 \times 6240 = 24,960$ temporal modes.

applying the van Loock–Furusawa criterion, we verified the 2D entanglement structure of the state if we observed that the variances of all the nullifiers were below -4.5 dB compared with shot noise (17).

The measurement results for each nullifier type for each temporal mode index k are shown in Fig. 4. All nullifiers were observed to be below -4.5 dB for up to $k = 6240$. Because there are four micromodes at each temporal index k , the states we verified possess 24,960 micromodes. Because one macromode (or one site) consists of four micromodes and we used $N = 5$, the structure of the state we verified is a 5- by 1248-site 2D square lattice with cylindrical boundary conditions. The means of the variances for each type of nullifier are -4.82 ± 0.06 dB, -5.34 ± 0.06 dB, -4.81 ± 0.06 dB, and -4.93 ± 0.06 dB, respectively. These values are limited by the original squeezing level from the squeezed light sources, optical losses and fluctuations in the optical system, and electrical noises from homodyne detectors and are in good agreement with the experimental parameters. Statistical errors are the main contributors to the error bars, which can be arbitrarily decreased by increasing the number of events used for calculating the nullifiers. There are no corrections for experimental imperfections, and the nullifiers do not degrade with the increasing k , suggesting that k can be arbitrarily large.

Thus, we have proposed and verified the generation of a universal resource state for MBQC. Using this cluster state for large-scale MBQC requires a few additional steps. First, because the delay line ratio $N = 5$ sets the number of inputs in the effective quantum circuit, this must be increased. Second, because finite squeezing level sets the noise level when cluster states are used in MBQC (20), the squeezing level must be improved to be above the fault-tolerant threshold (22). It is possible to increase the delay line ratio by increasing the bandwidth of the squeezed light source, which

reduces wave packet size, and development of a low-loss optical delay line. Even with currently available technology (26, 27), we expect a number of the input modes on the order of 10^4 to be achievable, and if we consider the ultimate limit set by coherence time of the light source, the number of the input modes could be potentially increased to $\sim 10^{10}$ modes by use of a narrow-linewidth laser (28). Regarding improvement of the squeezing level to above the fault-tolerant threshold, the squeezing level of our cluster state can potentially reach -15 dB with the state-of-the-art squeezed light source (29), which begins to be within reach of the known thresholds (-15 to -17 dB) for fault tolerance in this architecture by using particular quantum-error-correcting codes (30). Moreover, it has recently been shown that it is possible to further relax the threshold to about -10 dB with analog quantum error correction and postselection (23). Therefore, we believe that demonstration of our cluster state provides a feasible way toward realization of a practical quantum computer.

REFERENCES AND NOTES

- M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
- J. M. Gambetta, J. M. Chow, M. Steffen, *npj Quantum Inf.* **3**, 2 (2017).
- K. R. Brown, J. Kim, C. Monroe, *npj Quantum Inf.* **2**, 16034 (2016).
- F. A. Zwanenburg et al., *Rev. Mod. Phys.* **85**, 961–1019 (2013).
- M. Veldhorst, H. G. J. Eenink, C. H. Yang, A. S. Dzurak, *Nat. Commun.* **8**, 1766 (2017).
- H. J. Briegel, R. Raussendorf, *Phys. Rev. Lett.* **86**, 910–913 (2001).
- R. Raussendorf, H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
- N. C. Menicucci et al., *Phys. Rev. Lett.* **97**, 110501 (2006).
- N. C. Menicucci, *Phys. Rev. A* **83**, 062314 (2011).
- R. N. Alexander, S. Yokoyama, A. Furusawa, N. C. Menicucci, *Phys. Rev. A* **97**, 032302 (2018).
- R. N. Alexander et al., *Phys. Rev. A* **94**, 032327 (2016).
- A. L. Grimsmo, A. Blais, *npj Quantum Inf.* **3**, 20 (2017).
- M. Yukawa, R. Ukai, P. van Loock, A. Furusawa, *Phys. Rev. A* **78**, 012301 (2008).
- J. Yoshikawa et al., *APL Photonics* **1**, 060801 (2016).
- S. Yokoyama et al., *Nat. Photonics* **7**, 982–986 (2013).
- M. Chen, N. C. Menicucci, O. Pfister, *Phys. Rev. Lett.* **112**, 120505 (2014).
- Materials and methods are available as supplementary materials.

- Y. Shiozawa et al., *Phys. Rev. A* **98**, 052311 (2018).
- S. Lloyd, S. L. Braunstein, *Phys. Rev. Lett.* **82**, 1784–1787 (1999).
- R. N. Alexander, S. C. Armstrong, R. Ukai, N. C. Menicucci, *Phys. Rev. A* **90**, 062324 (2014).
- M. V. Larsen, X. Guo, C. R. Breum, J. S. Neergaard-Nielsen, U. L. Andersen, *Science* **366**, 369–372 (2019).
- N. C. Menicucci, *Phys. Rev. Lett.* **112**, 120504 (2014).
- K. Fukui, A. Tomita, A. Okamoto, K. Fujii, *Phys. Rev. X* **8**, 021054 (2018).
- D. Gottesman, A. Kitaev, J. Preskill, *Phys. Rev. A* **64**, 012310 (2001).
- B. Q. Baragiola, G. Pantaleoni, R. N. Alexander, A. Karanjai, N. C. Menicucci, All-Gaussian universality and fault tolerance with the Gottesman-Kitaev-Preskill code. arXiv:1903.00012 [quant-ph] (2019).
- S. Ast, M. Mehmet, R. Schnabel, *Opt. Express* **21**, 13572–13579 (2013).
- J. Aasi et al., *Nat. Photonics* **7**, 613–619 (2013).
- L. Wu et al., *Sci. Rep.* **6**, 24969 (2016).
- H. Vahlbruch, M. Mehmet, K. Danzmann, R. Schnabel, *Phys. Rev. Lett.* **117**, 110801 (2016).
- B. W. Walshe, L. J. Mensen, B. Q. Baragiola, N. C. Menicucci, *Phys. Rev. A* **100**, 010301 (2019).

ACKNOWLEDGMENTS

Funding: This work was partly supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant 18H05207), UTokyo Foundation, donations from Nichia Corporation, and the Australian Research Council Centre of Excellence for Quantum Computation and Communication Technology (project CE170100012). W.A. acknowledges financial support from JSPS. Y.S. acknowledges financial support from the Advanced Leading Graduate Course for Photon Science (ALPS). B.C. acknowledges financial support from the Program of Excellence in Photon Science (XPS). R.N.A. is supported by National Science Foundation grant PHY-1630114.

Author contributions: W.A. and Y.S. planned and designed the experiment. S.Y., S.T., J.Y., H.Y., and A.F. supervised this project. R.N.A., N.C.M., W.A., and S.Y. formulated the theory for this experiment. Y.S. and W.A. designed the actual optical system. W.A. conceived the method for stabilizing the experimental system and performed analysis necessary for experimental realization. W.A., Y.S., and H.E. built the experimental system. W.A., Y.S., and B.C. conducted the experiment and obtained the experimental data. Y.S., B.C., W.A., and S.Y. performed the data analysis. W.A. wrote the manuscript with assistance from Y.S., B.C., S.Y., R.N.A., S.T., J.Y., N.C.M., H.Y., and A.F. **Competing financial interests:** The authors declare no competing financial interests.

Data and materials availability: All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/373/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
References (31–37)

4 June 2019; accepted 25 September 2019
10.1126/science.aay2645

NEUROSCIENCE

Isolated cortical computations during delta waves support memory consolidation

Ralitsa Todorova and Michaël Zugaro*

Delta waves have been described as periods of generalized silence across the cortex, and their alternation with periods of endogenous activity results in the slow oscillation of slow-wave sleep. Despite evidence that delta waves are instrumental for memory consolidation, their specific role in reshaping cortical functional circuits remains puzzling. In a rat model, we found that delta waves are not periods of complete silence and that the residual activity is not mere neuronal noise. Instead, cortical cells involved in learning a spatial memory task subsequently formed cell assemblies during delta waves in response to transient reactivation of hippocampal ensembles during ripples. This process occurred selectively during endogenous or induced memory consolidation. Thus, delta waves represent isolated cortical computations tightly related to ongoing information processing underlying memory consolidation.

Most of our time spent asleep is dominated by slow oscillations (0.1 to 1 Hz), when cortical neurons synchronously alternate between a depolarized (up) state associated with high levels of endogenous activity and a hyperpolarized (down) state when neurons remain silent (1). Delta waves are large deflections of the local field potential (LFP) that correspond to the down states of the slow oscillation

and are thus considered periods of generalized cortical silence. The slow oscillation plays a causal role in memory consolidation (2–5), in particular by orchestrating an information flow between the hippocampus and the neocortex (6). Indeed, delta waves tend to occur in close temporal proximity to hippocampal ripples (7), which are instrumental for memory consolidation (8, 9). Hippocampal replay of awake activity (10), biased by

inputs from sensory cortices (11, 12), initiates reactivation of prefrontal cortical cell assemblies (13, 14) just before the occurrence of a delta wave (7). Cortical synaptic plasticity subsequently takes place during network reorganization early in the following up state (15, 16) and during the massive calcium entry accompanying the ensuing sleep spindle (17–19). This hippocampo-cortical dialogue (20–22) is instrumental for memory consolidation (5). However, the incursion of generalized silence (delta wave) precisely between periods of information exchange and periods of network plasticity remains puzzling.

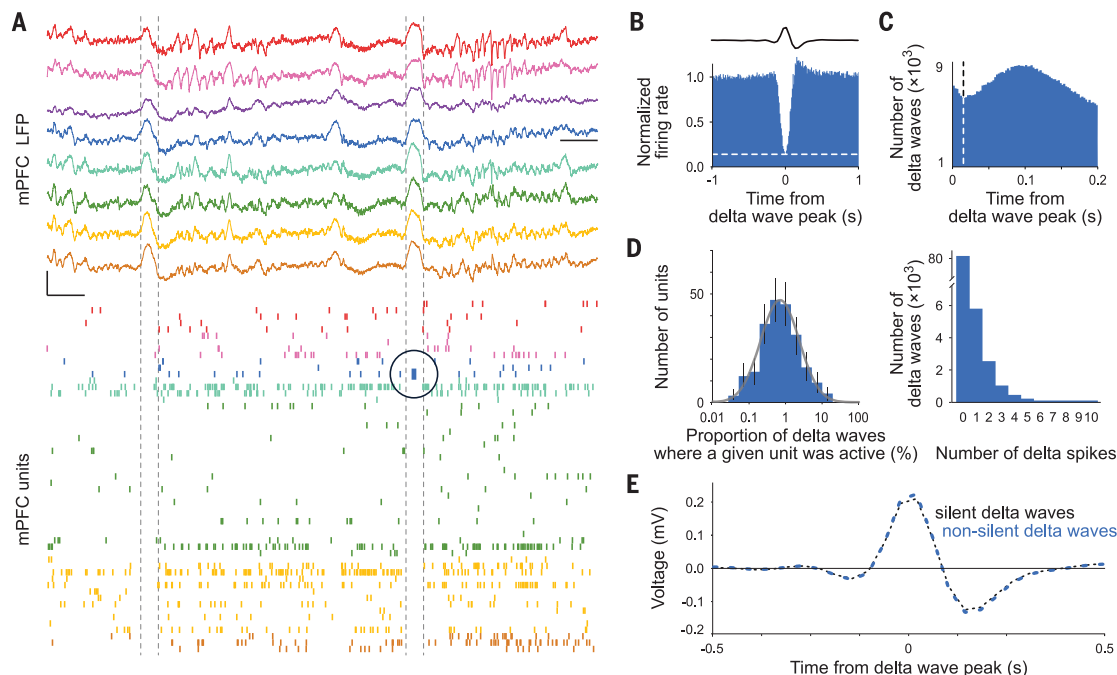
We recorded prefrontal cortical activity in nine rats during slow-wave sleep (5). Consistent with previous reports, most delta waves were accompanied by neuronal silence. Yet, occasionally, spikes did occur during delta waves (Fig. 1A), and when considering cumulative spiking activity over all recorded delta waves, unexpected residual activity appeared at the peak of the waves (Fig. 1, B and C) (spike waveforms recorded during delta waves were not distinguishable from spike waveforms

Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, Université PSL, Paris, France.

*Corresponding author. Email: michael.zugaro@college-de-france.fr

Fig. 1. Delta spikes.

(A) Example of a nonsilent delta wave. Colored curves indicate LFPs recorded from the medial prefrontal cortex (color: recording channel). Colored vertical ticks indicate spikes emitted by simultaneously recorded prefrontal units (color: channel from which the unit was recorded). Dashed lines indicate the beginning and end of delta waves. A delta spike (black circle) occurs during the second delta wave, when the rest of the network remains silent. Black calibration bars: 0.5 s (horizontal); 1 mV (vertical). mPFC, medial prefrontal cortex. (B) Mean perievent time histogram of the normalized firing rate of prefrontal units centered on delta waves (top curve: mean field event). The dashed white line indicates residual activity during delta waves. (C) Time distribution of the spikes emitted by each prefrontal neuron closest to each delta wave. The large peak at ~100 ms corresponds with activity in the up state. The smaller peak consists of spikes occurring during delta waves. The dashed line indicates the 15-ms upper threshold used to define delta spikes in subsequent analyses (all results were confirmed by using ± 30 -ms time windows).



(D) (Left) Number of units that discharged in a given proportion of delta waves (gray curve: log-normal fit with the same mean and variance as the data; error bars: 95% confidence intervals from bootstrapped data). No unit fired in 0% of the delta waves. (Right) Number of delta waves containing a given number of delta spikes. (E) No difference in average waveforms between silent (black: $n = 101,161$) and nonsilent (blue: $n = 12,205$) delta waves (Monte Carlo test, $P > 0.05$).

recorded outside delta waves; fig. S1). On closer examination, neuronal activity occurred consistently in a substantial fraction of delta waves (12%), wherein one or a few neurons remained active while the rest of the population became silent (Fig. 1D). We call this unexpected persisting activity “delta spikes.” To investigate whether delta spikes were restricted to a particular subset of neurons, we counted the number of delta waves in which each unit emitted one or more spikes. As it happened, every single recorded unit fired during delta waves, suggesting instead that persisting firing may actually constitute a widespread phenomenon (Fig. 1D and fig. S2).

We then wondered whether delta spikes tended to occur in specific delta waves with

distinctive characteristics. We thus compared delta waves in which we did or did not detect cortical spikes and found no significant difference between the two groups in terms of waveform (Fig. 1E), duration, timing (fig. S3), depth (fig. S4), decreased gamma power, or coupling with hippocampal ripples and thalamocortical spindles (fig. S5). This suggests that spikes could take place during virtually all delta waves but may remain undetected given the limited number of recorded neurons relative to the entire population (fig. S6). We thus hypothesize that firing during delta waves might be an overlooked phenomenon manifested in possibly all delta waves.

These findings indicate that during any given delta wave, the cortical network becomes

silent except for a small but ever-changing minority of cells. The most parsimonious explanation would be that delta spikes constitute random activity reflecting imperfect coordination in the cortical alternation between the up and down states. Yet, an alternative possibility is that this activity serves a well-defined computational function. A hallmark of cortical computation is the emergence of cell assemblies. We thus tested for the presence of recurring coactive cell ensembles, using two complementary approaches. As a first approach, we performed a standard independent component analysis (23), which identified multiple significant components that were active during delta waves (fig. S7, A and B). However, these components were likely to

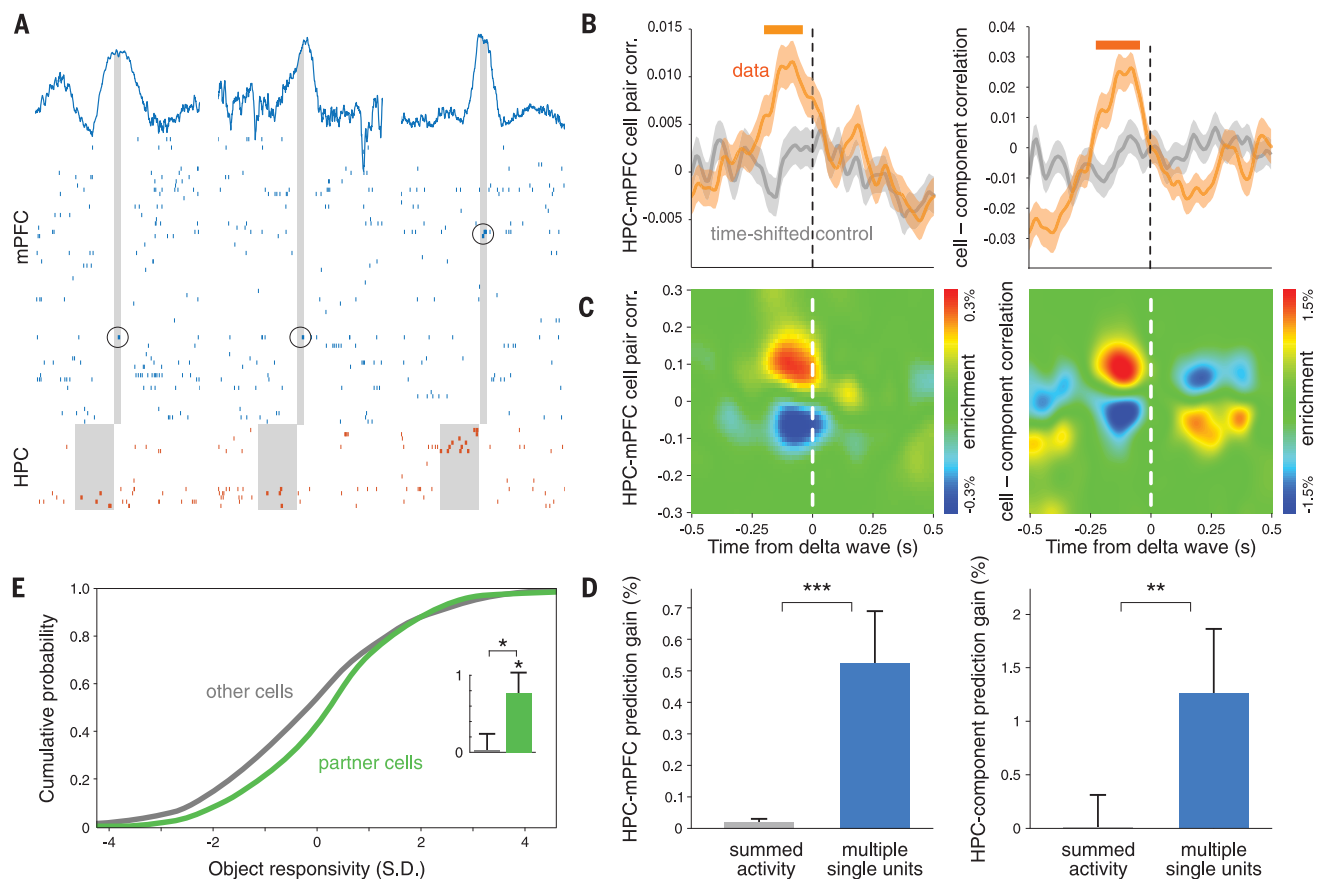


Fig. 2. Hippocampal ripple activity predicts delta spikes. (A) Delta spikes and preceding hippocampal activity. Blue traces indicate LFPs from the medial prefrontal cortex. Colored ticks indicate simultaneously recorded prefrontal (blue) and hippocampal (red) spikes. Black circles indicate delta spikes emitted within ± 15 ms of the delta wave peak (shaded area). In the first two delta waves, delta spikes were recorded from the same unit after similar hippocampal activity patterns. HPC, hippocampus. (B) Cross-correlations (curves and shaded areas, mean \pm SEM; orange, data; gray, time-shifted control) between hippocampal ripple activity (sliding window) and delta spikes (fixed, 0 s). The horizontal orange line indicates the Monte Carlo test: $P < 0.05$. corr., correlation. (C) Enrichment in positive correlations [comparative distribution between data and control in (B)] when hippocampal activity was correlated with subsequent prefrontal

delta spikes. (D) Performance of a GLM trained to predict prefrontal activity during delta waves on the basis of preceding hippocampal ripple activity (200-ms window). Delta spikes as well as delta components were significantly predicted by multiple single-unit hippocampal activity ($P = 0.0403$ for delta spikes, $P = 0.0052$ for delta components; Wilcoxon rank sum test) but not by global hippocampal drive ignoring cell identity (summed activity, $P = 0.2597$ for delta spikes and $P = 0.3258$ for delta components; Wilcoxon rank sum test). $**P < 0.01$; $***P < 0.001$, (Wilcoxon signed-rank tests). (E) Object responsiveness index for partner (green) versus other (gray) prefrontal units (curves: cumulative distributions; inset: mean \pm SEM). Only partner prefrontal units showed positive object responsiveness (partner units, $P = 0.0162$; other units, $P = 0.5967$; Wilcoxon signed-rank test; partner versus other units, $P = 0.0465$; Wilcoxon rank sum test). $*P < 0.05$.

combine multiple smaller but overlapping cell ensembles, given the limited number of neurons active in any given delta wave. We thus performed a second analysis to examine cooperative activity (“peer prediction”) (24) among delta spikes, an idiosyncratic property of cell assemblies. This showed that the delta spikes of one neuron could be predicted from the delta spikes of other neurons (fig. S7C).

We then asked whether delta spikes were involved in the hippocampo-cortical dialogue underlying memory consolidation. Because delta waves typically take place precisely between hippocampal replay and cortical reorganization for memory consolidation, this hypothesis would be expected to have two implications: (i) Hippocampal activity during ripples should predict which neurons (or which

assemblies) are active during the following delta wave, and (ii) this predictive bias should emerge after behavior, and predictable cortical cells should be involved in the reactivation of waking experience.

Rats were trained on a spatial memory task, and hippocampal and cortical activity was recorded during both behavior and subsequent memory consolidation during the first 2 hours

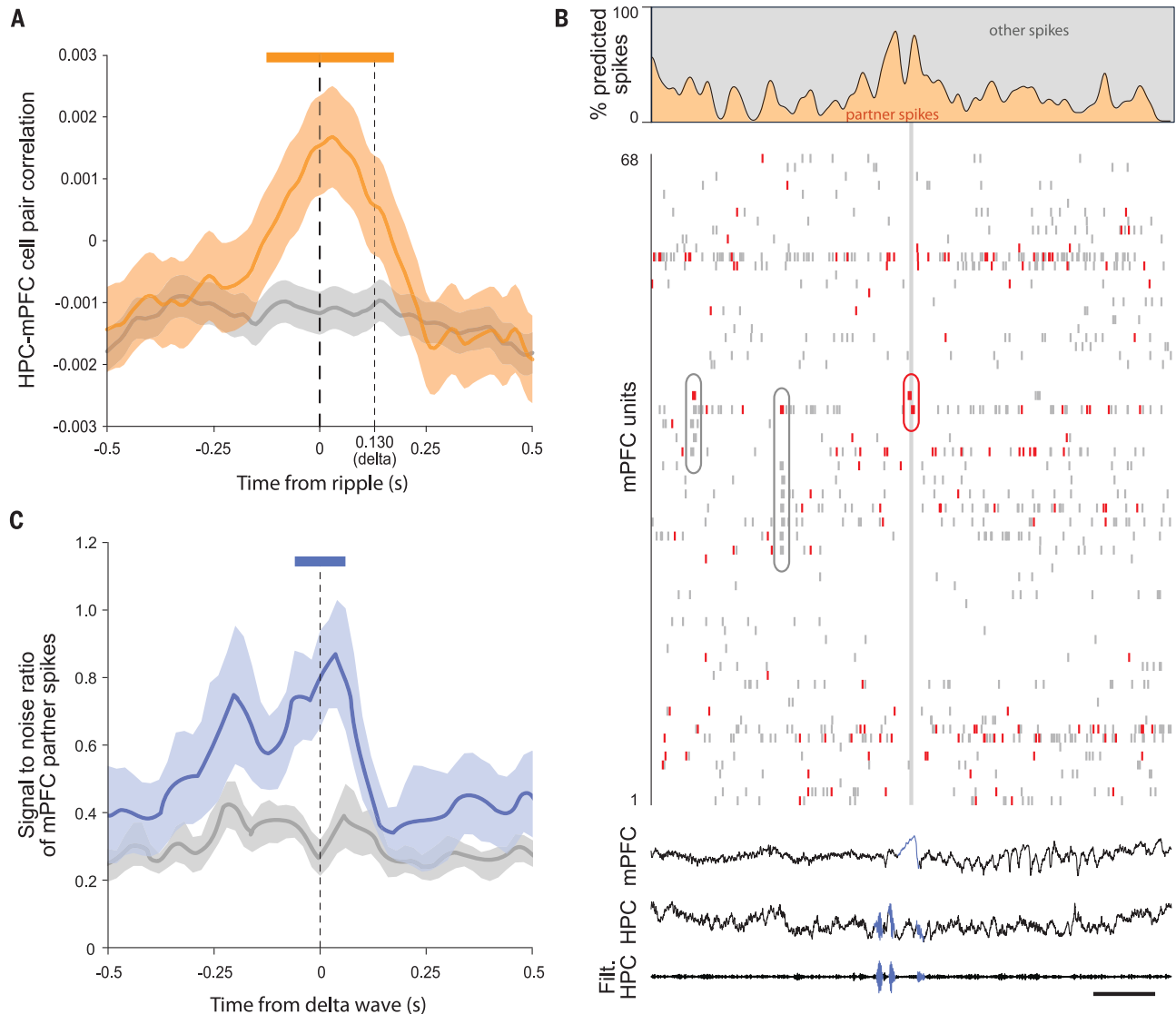


Fig. 3. Delta waves isolate cortical computations. (A) Cross-correlations (curves and shaded areas, mean \pm SEM) between hippocampal ripple activity (fixed, 0 s) and prefrontal activity (sliding window). Observed cross-correlations (orange) were significantly different from a time-shifted control (gray) for cortical activity after ripples (horizontal orange line: Monte Carlo test, $P < 0.05$). Delta waves (dashed line; peak occurrence rate 130 ms after ripples) tend to occur within the critical window when prefrontal activity remains correlated with the preceding ripple activity. (B) Simultaneous recording of prefrontal and hippocampal activity around a delta wave (gray-shaded rectangle). (Top) Proportion of prefrontal spikes predicted by the firing of hippocampal cells (partner spikes). (Center) Raster plot of spikes

emitted by 68 simultaneously recorded prefrontal units (red ticks: partner spikes; gray ticks: other spikes). (Bottom) Simultaneously recorded LFPs in the mPFC (blue: delta wave) and hippocampus (broadband and ripple-band filtered signal; blue: ripples). During delta waves, partner spikes occurred in isolation (red ellipse). Partner spikes emitted by the same units outside delta waves (gray ellipses) formed a considerably smaller proportion of the ongoing cortical activity. Black calibration bar: 0.5 s. Filt., filtered. (C) Signal-to-noise ratio (curves and shaded areas, mean \pm SEM) of partner spikes relative to other spikes around delta waves. Observed values (blue) were significantly different from a time-shifted control (gray) during delta waves (horizontal blue line: Monte Carlo test, $P < 0.05$).

of sleep (5). Hippocampal spiking activity during ripples was significantly correlated with cortical delta spikes that occurred immediately (50 to 200 ms) afterward (the effect was not due to data recorded from any single

rat) (Fig. 2, A and B, and fig. S8). This increased correlation was due to a large proportion of positively correlated interregional pairs of neurons (Fig. 2C). Furthermore, ripple spikes were better correlated with delta spikes than

with spikes occurring at similar delays during an up state (correlations were greater in the presence of a delta wave) (fig. S9). A generalized linear model (GLM) analysis showed that ensemble activity in the hippocampus

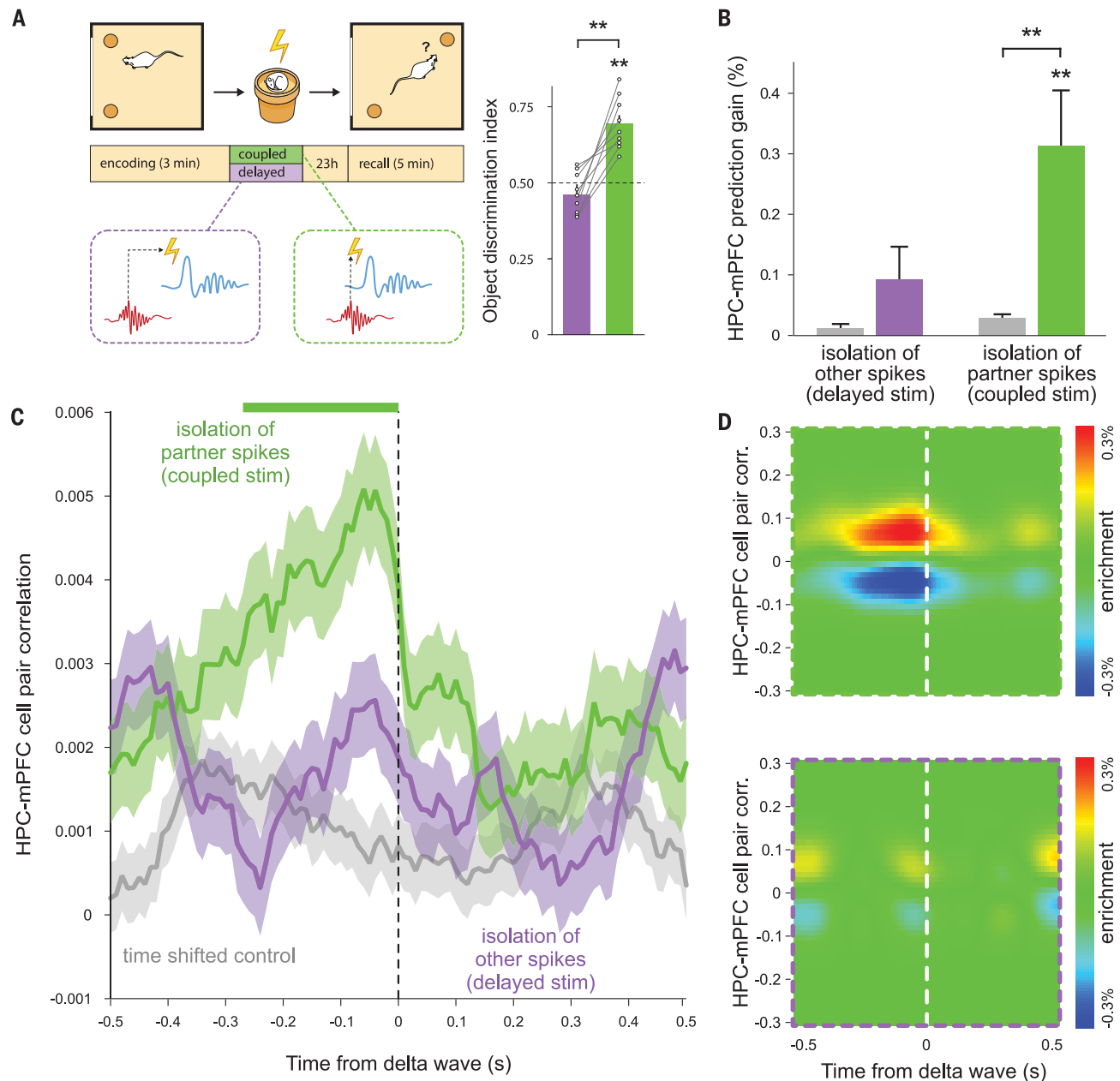


Fig. 4. Induction of memory consolidation by isolation of partner spikes.

(A) (Left) Experimental protocol. Delta waves were induced by brief single-pulse electrical stimulation of deep cortical layers. Induced delta waves were triggered to isolate either partner activity (coupled stimulation; green, 130 ms after ripples) or other cortical activity (delayed stimulation; purple, 290 to 370 ms after ripples) during sleep after limited training on a spatial object-recognition task. (Right) Object discrimination index during the recall phase. Only delta waves triggered to isolate partner activity (coupled stimulation) resulted in memory consolidation and enhanced task performance. (B) Performance of a GLM trained to predict delta spikes on the basis of preceding hippocampal activity (200-ms window), measured as

percent improvement relative to a shuffled control (prediction gain). Only delta waves triggered to isolate partner activity resulted in a significant prediction of delta spikes ($P = 0.0030$ for isolation of partner spikes by coupled stimulation; $P = 0.1301$ for isolation of other spikes by delayed stimulation; Wilcoxon rank sum test). stim, stimulation. $**P < 0.01$. (C) Cross-correlation (curves and shaded areas, mean \pm SEM) of hippocampal activity and delta spikes (green: isolation of partner spikes by coupled stimulation; purple: isolation of other spikes by delayed stimulation; gray: time-shifted control; horizontal green line: Monte Carlo test, $P < 0.05$). (D) Enrichment in positive correlations upon isolation of partner spikes by coupled stimulation (top) but not of other spikes by delayed stimulation (bottom).

could significantly predict which cortical cells would emit delta spikes (Fig. 2D). In contrast, delta spikes could not be predicted from the combined activity of all hippocampal units that ignored cell identity (multiunit activity), ruling out the possibility that delta spikes merely reflect the overall level of hippocampal excitatory drive during ripples. Finally, ripples facilitated (but did not entirely control) the formation of delta assemblies (fig. S7). Furthermore, the same GLM analysis applied to hippocampal and cortical ensembles showed that hippocampal activity could even predict delta components (Fig. 2D).

Our second prediction concerned the relation of this predictive bias to behavior. In sleep sessions preceding the task, hippocampal spiking activity during ripples failed to predict subsequent delta spikes or assemblies (fig. S10), indicating that the predictive bias emerged after task performance. We then investigated the behavioral correlates of the prefrontal units whose delta spikes were significantly predicted by hippocampal ripple activity during sleep after behavior (“partner cells”) (fig. S11 and tables S1 and S2). These cortical cells displayed higher levels of task-relevant firing during behavior (Fig. 2E) (we failed to find a similar effect for delta components, possibly because of low statistical power due to their limited number: $n = 14$ components, $n = 9$ predicted). We further investigated the behavioral correlates of delta assemblies and found that these assemblies were also expressed during task performance but not outside delta waves nor in sleep that preceded behavior (fig. S12).

These observations suggest that in addition to triggering the reorganization of cortical subnetworks during the transition to the up state (5), an unsuspected role of the delta wave may be to isolate from interference specific cortical computations taking place in response to hippocampal replay. Consistent with this idea, delta waves typically occurred within a critical time window when cortical activity remained correlated with hippocampal ripple activity (Fig. 3A). To test whether delta waves tended to preferentially silence cortical activity that was related to the ongoing hippocampo-cortical dialogue, we classified individual prefrontal spikes as “partner spikes” if they followed spikes emitted by their significantly correlated hippocampal units or “other spikes” if they were unrelated to the preceding hippocampal activity (Fig. 3B) (more examples are shown in fig. S13). The signal-to-noise ratio for partner spikes peaked during delta waves (Fig. 3C), and this was due to the selective silencing of non-partner activity during delta waves (fig. S14).

Does this isolation of cortical computations play a critical role in memory consolidation? A prediction of this hypothesis is that isolating cortical assemblies by experimental induction of delta waves should trigger memory consolidation, but only if the isolated activity is relevant to the hippocampo-cortical dialogue (partner spikes). We have already shown that triggering delta waves when endogenous mechanisms fail to do so can boost memory consolidation, provided that the delta waves are induced in an appropriate time window (Fig. 4A) (5). We thus sought to confirm the prediction that these delta waves actually isolated partner spikes (that delta spikes did occur during induced delta waves and that they were predicted by hippocampal activity). Similar to our observations in natural sleep (above), stimulation-induced delta waves did feature spiking activity (fig. S15), and these delta spikes were predicted by preceding hippocampal activity coinciding with the timing of ripples (Fig. 4, B to D). In contrast, slightly delaying the induction of delta waves (by ~200 ms) (Fig. 3A) to isolate nonpartner delta spikes (Fig. 4, B to D) failed to induce memory consolidation (5).

Our results challenge the generally accepted tenet that delta waves, reflecting the down states of the sleep slow oscillation, are periods of complete cortical silence (1, 6, 25), to the point that they have sometimes been defined as such (26, 27) and that occasional spikes have been routinely ignored when detected (28, 29). We focused on delta spikes and found that they are not neuronal noise due to imperfect silencing of the cortical mantle. On the contrary, they constitute a common phenomenon potentially implicating all neurons and all delta waves, and they reflect genuine processing involved in memory consolidation.

This observation also provides a mechanism for the documented but puzzling role of delta waves in memory consolidation: Synchronized silence across most of the cortex isolates the network from competing inputs while a select subpopulation of neurons maintains relevant spike patterns between epochs of hippocampo-cortical information transfer (10, 12, 14) and epochs of cortical plasticity (15, 16) and network reorganization (5, 18, 19). Yet, in many cases, cortical activity during delta waves could not be reliably predicted from the preceding hippocampal ripple activity. Such cortical activity could instead have been related to interactions with other brain networks. This suggests that delta spikes and assemblies might constitute a general mechanism of isolated cortical computation beyond the hippocampo-cortical dialogue.

REFERENCES AND NOTES

1. M. Steriade, A. Nuñez, F. Amzica, *J. Neurosci.* **13**, 3252–3265 (1993).
2. H.-V. V. Ngo, T. Martinetz, J. Born, M. Mölle, *Neuron* **78**, 545–553 (2013).
3. L. Marshall, H. Helgadottir, M. Mölle, J. Born, *Nature* **444**, 610–613 (2006).
4. S. Chauvette, J. Seigne, I. Timofeev, *Neuron* **75**, 1105–1113 (2012).
5. N. Maingret, G. Girardeau, R. Todorova, M. Goutier, M. Zugaro, *Nat. Neurosci.* **19**, 959–964 (2016).
6. A. Sirota, G. Buzsáki, *Thalamus Relat. Syst.* **3**, 245–259 (2005).
7. A. Sirota, J. Csicsvari, D. Buhl, G. Buzsáki, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2065–2069 (2003).
8. G. Girardeau, K. Benchenane, S. I. Wiener, G. Buzsáki, M. B. Zugaro, *Nat. Neurosci.* **12**, 1222–1223 (2009).
9. V. Ego-Stengel, M. A. Wilson, *Hippocampus* **20**, 1–10 (2010).
10. A. K. Lee, M. A. Wilson, *Neuron* **36**, 1183–1194 (2002).
11. D. Ji, M. A. Wilson, *Nat. Neurosci.* **10**, 100–107 (2007).
12. G. Rothschild, E. Eban, L. M. Frank, *Nat. Neurosci.* **20**, 251–259 (2017).
13. C. M. Wierzyński, E. V. Lubenov, M. Gu, A. G. Siapas, *Neuron* **61**, 587–596 (2009).
14. A. Peyrache, M. Khamassi, K. Benchenane, S. I. Wiener, F. P. Battaglia, *Nat. Neurosci.* **12**, 919–926 (2009).
15. P. B. Kruskal, L. Li, J. N. MacLean, *Nat. Commun.* **4**, 2574 (2013).
16. T. Gulati, L. Guo, D. S. Ramanathan, A. Bodepudi, K. Ganguly, *Nat. Neurosci.* **20**, 1277–1284 (2017).
17. A. G. Siapas, M. A. Wilson, *Neuron* **21**, 1123–1128 (1998).
18. M. Rosanova, D. Ulrich, *J. Neurosci.* **25**, 9398–9405 (2005).
19. J. Seibt et al., *Nat. Commun.* **8**, 684 (2017).
20. D. Marr, *Philos. Trans. R. Soc. London Ser. B* **262**, 23–81 (1971).
21. G. Buzsáki, *Neuroscience* **31**, 551–570 (1989).
22. P. W. Frankland, B. Bontempi, *Nat. Rev. Neurosci.* **6**, 119–130 (2005).
23. V. Lopes-dos Santos, S. Ribeiro, A. B. L. Tort, *J. Neurosci. Methods* **220**, 149–166 (2013).
24. K. D. Harris, J. Csicsvari, H. Hirase, G. Dragoi, G. Buzsáki, *Nature* **424**, 552–556 (2003).
25. M. Steriade, I. Timofeev, *Neuron* **37**, 563–576 (2003).
26. A. Luczak, P. Barthó, S. L. Marguet, G. Buzsáki, K. D. Harris, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 347–352 (2007).
27. L. A. Johnson, D. R. Euston, M. Tatsuno, B. L. McNaughton, *J. Neurosci.* **30**, 2650–2661 (2010).
28. J. Csicsvari et al., *J. Neurophysiol.* **90**, 1314–1323 (2003).
29. D. Jercog et al., *eLife* **6**, e22425 (2017).
30. N. Maingret, R. Todorova, M. Zugaro, Recordings and timed stimulation of rat dorsal hippocampal area CA1 and medial prefrontal cortex during behavior and sleep. CRCNS (2019); <https://doi.org/10.6080/KOTD9VJG>.

ACKNOWLEDGMENTS

We thank N. Maingret for collecting the data and R. Fayat for help with the delta wave detection algorithm. **Funding:** This work was supported by the Agence Nationale de la Recherche (ANR-15-CE16-0001-02 and ANR-17-CE37-0016-01), Collège de France (R.T.), and a joint grant from École des Neurosciences de Paris Île-de-France and LabEx MemoLife (ANR-10-LABX-54 MEMO LIFE and ANR-10-IDEX-0001-02 PSL*) (R.T.). **Author contributions:** M.Z. designed the study. R.T. and M.Z. designed the analyses and wrote the manuscript. R.T. performed the analyses. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data in this study are available at CRCNS (30).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/377/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 and S2
References (31–34)

[View/request a protocol for this paper from Bio-protocol.](#)

16 May 2019; accepted 10 September 2019
10.1126/science.aay0616

TURBULENCE

Synthetic dissipation and cascade fluxes in a turbulent quantum gas

Nir Navon^{1*}, Christoph Eigen², Jinyi Zhang², Raphael Lopes^{2†}, Alexander L. Gaunt^{2,3}, Kazuya Fujimoto^{4*‡}, Makoto Tsubota⁵, Robert P. Smith^{2,6}, Zoran Hadzibabic²

Scale-invariant fluxes are the defining property of turbulent cascades, but their direct measurement is a challenging experimental problem. Here we perform such a measurement for a direct energy cascade in a turbulent quantum gas. Using a time-periodic force, we inject energy at a large length scale and generate a cascade in a uniformly trapped three-dimensional Bose gas. The adjustable trap depth provides a high-momentum cutoff k_D , which realizes a synthetic dissipation scale. This gives us direct access to the particle flux across a momentum shell of radius k_D , and the tunability of k_D allows for a clear demonstration of the zeroth law of turbulence. Moreover, our time-resolved measurements give unique access to the pre-steady-state dynamics, when the cascade front propagates in momentum space.

The discovery of a universal law describing the transfer of energy from large to small length scales in turbulent flows was a conceptual breakthrough (1, 2). Despite their complex spatiotemporal dynamics, turbulent flows often obey a simple generic picture: The energy injected into the system at a large length scale is gradually transferred to ever smaller ones, flowing locally in Fourier space through the so-called inertial range where no dissipation occurs, until it is dissipated at some small length scale. In Fig. 1A, we depict such turbulent-cascade dynamics for a compressible field in real space. The field is initially at rest. At times $t > 0$, an external force creates excitations at a large length scale $1/k_F$. These excitations propagate to smaller length scales owing to their nonlinear interactions. Once they first reach the dissipation scale $1/k_D$, at time t_d , the field fluctuates on all length scales from $1/k_F$ to $1/k_D$. If a steady state is established within the momentum range k_F to k_D , from thereon energy is dissipated at k_D at the same rate at which it is injected at k_F . In such a steady state, the momentum-space distributions of quantities such as the energy or wave amplitude are generically scale-free power laws.

Many quantitative theoretical predictions about turbulence are based on taking the mathematical limits $k_F \rightarrow 0$ and $k_D \rightarrow \infty$ (3).

Such formal treatments lead to predictions that are elegant but often counterintuitive. A key prediction of this kind is that for $k_D \rightarrow \infty$, the steady-state cascade corresponds to a scale-invariant (k -independent) energy flux through momentum space, but no particle flux (4).

Experimentally, the steady-state power-law spectra of various quantities have been extensively studied (5–9), whereas the equally fundamental cascade fluxes are harder to measure (10–13). Recently, ultracold atomic gases have emerged as a versatile platform for studies of turbulence (9, 14–22), offering experimental possibilities unavailable in other systems. Here, we use an atomic gas to directly measure cascade fluxes in a turbulent system. Our dissipation scale is tuneable, which allows us to explore how the fluxes depend on k_D , and to reconcile the experimental observations with the formal predictions for $k_D \rightarrow \infty$. Our system also allows a time-resolved study of the initial stage of turbulence (23–25), when a steady state is not yet established, which reveals how the cascade front propagates in momentum space.

Our experiment starts with a weakly interacting Bose-Einstein condensate of $N \approx 1.2 \times 10^5$ atoms of ^{87}Rb in the uniform potential of a cylindrical optical box trap of radius $R \approx 16 \mu\text{m}$ and length $L \approx 27 \mu\text{m}$ (Fig. 1B) (26). At the end of the initial preparation of the gas, the noncondensed fraction is $<10\%$ and the chemical potential is $\mu \approx k_B \times 2 \text{ nK}$, corresponding to a healing length $\xi \approx 1.2 \mu\text{m} \ll R, L$. We initiate a turbulent cascade by injecting energy at the system-size length scale (corresponding to a small momentum k_F), using a spatially uniform force $\mathbf{F}_s(\mathbf{r}, t) = F_0 \sin(\omega_s t) \hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is a unit vector along the box symmetry axis, $F_0 L \approx k_B \times 2.5 \text{ nK}$, and $\omega_s \approx 2\pi \times 9 \text{ Hz}$ is tuned to resonantly excite the sound wave of wavelength $2L$ (so $k_F = \pi/L$) (27). This anisotropic forcing of the matter-wave field is represented in Fig. 1C as a small dark blue

area elongated along k_x . As shown in (9), after several seconds of shaking, the momentum distribution of the gas in the inertial range is statistically isotropic and time-invariant, $n(\mathbf{k}, t) \approx \langle n(k) \rangle \propto k^{-\gamma}$, with $\gamma \approx 3.5$ (28, 29). The time invariance implies that the energy and particle fluxes through this k -range are k -independent, but it does not reveal their values. Here, we extract the cascade fluxes by studying the dissipation in our gas.

In conventional fluids, one observes macroscopic (hydrodynamic) degrees of freedom, and the dissipation occurs in the form of heating (i.e., transfer of energy into the microscopic degrees of freedom). This dissipation is set by the viscosity ν , which is generally not tuneable. Moreover, the resulting minute heating is often difficult to measure because of thermal coupling of the fluid with its surroundings (30). Our system is thermally isolated from the environment, and we have direct access to all the microscopic degrees of freedom, so the dissipation occurs only in the form of (readily measurable) particle loss. The optical box (Fig. 1B) has a non-infinite energy depth U_D , so particles with a sufficiently large energy leave the box; in momentum space, U_D corresponds to a sphere of radius $k_D = \sqrt{2mU_D}/\hbar$ (Fig. 1C), where m is the atom mass and \hbar is Planck's constant divided by 2π . This simple feature realizes a synthetic dissipation scale, with U_D defining the particle and energy sink. Crucially, this dissipation scale can be tuned by changing the trapping laser power (31).

Formally, within the assumptions of the wave-turbulence theory, the equations of motion lead to a continuity equation, with a source and a sink, that is local in momentum space (3):

$$\frac{\partial n(\mathbf{k}, t)}{\partial t} = F(\mathbf{k}, t) - D(\mathbf{k}, t) - \nabla_{\mathbf{k}} \cdot \Pi_n(\mathbf{k}, t) \quad (1)$$

Here $F(\mathbf{k}, t)$ corresponds to the external force, $D(\mathbf{k}, t)$ describes the dissipation, and $\nabla_{\mathbf{k}} \cdot \Pi_n$ captures the nonlinear interactions, where Π_n is the particle flux. For $F = D = 0$, the steady-state solutions are zero-flux equilibrium thermodynamic states. If F and D are nonzero but are localized in k space, one can also get non-equilibrium steady-state solutions with a nonzero scale-independent flux sustained by the source F and the sink D .

For an isotropic outflow, the total radial particle flux is $\Pi_n(k) = 4\pi k^2 |\Pi_n(\mathbf{k})|$. Hence, from Eq. 1, in the inertial range $4\pi k^2 \partial n / \partial t = -\partial \Pi_n / \partial k$. Integrating over k shows that we can measure the particle flux through the shell at k_D by simply counting the atoms remaining in the trap (see Fig. 1C):

$$\frac{\partial N}{\partial t} \equiv -\Pi_n(k_D, t) \quad (2)$$

¹Department of Physics, Yale University, New Haven, CT 06520, USA. ²Cambridge Laboratory, University of Cambridge, Cambridge CB3 0HE, UK. ³Microsoft Research, Cambridge CB1 2FB, UK. ⁴Department of Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ⁵Department of Physics and Nambu Yoichiro Institute of Theoretical and Experimental Physics, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi-Ku, Osaka 558-8585, Japan. ⁶Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, UK.

*Corresponding author. Email: nir.navon@yale.edu (N.N.); kazuya_fujimoto@rover.nuap.nagoya-u.ac.jp (K.F.) †Present address: Laboratoire Kastler Brossel, Collège de France, CNRS, ENS-PSL University, UPMC-Sorbonne Université, F-75005 Paris, France. ‡Present address: Institute for Advanced Research, Nagoya University, Nagoya 464-8601, Japan.

For a (non-equilibrium) steady state, with time-invariant $n(k)$ in the inertial range (9), the particle flux is k - and t -independent (32), so $\Pi_n(k_D, t) = \Pi_n(k, t) = \Pi_n$.

In steady state, the total radial energy flux, $\Pi_\varepsilon(k, t)$, is also k - and t -independent in the inertial range, and is equal to the rate of energy dissipation. To relate it to Π_n , we consider the pertinent case of weakly interacting particles with a dispersion relation $\omega(k)$, so the energy spectrum is $\mathcal{E}(k, t) = \hbar\omega(k)n(k, t)$; in our case, $\omega(k) \propto k^2$. At $k < k_D$, microscopic interactions drive particles to both lower and higher k , so the relationship between the net energy and particle fluxes, Π_ε and Π_n , is non-trivial; one might naively expect that $\Pi_\varepsilon(k) = \hbar\omega(k)\Pi_n(k)$, but this cannot be true if both Π_ε and Π_n are k -independent and $\omega(k)$ is not. However, at k_D the particles flow only one way because there is no “backflow” from the sink into the inertial range, so one can intuitively write

$$\Pi_\varepsilon(k_D) = \hbar\omega(k_D) \Pi_n(k_D) \quad (3)$$

Steady state then requires $\Pi_\varepsilon = \hbar\omega(k_D)\Pi_n$ at all k in the inertial range; for our $\omega(k)$, this means that $\Pi_\varepsilon \propto k_D^2 \Pi_n$. To formally derive Eq. 3, one multiplies Eq. 1 by $\hbar\omega(k)$ and invokes the continuity equation for the energy to obtain

$$\frac{\partial \Pi_\varepsilon(k, t)}{\partial k} = \hbar\omega(k) \frac{\partial \Pi_n(k, t)}{\partial k} \quad (4)$$

in the inertial range. For $k < k_D$, this equation is trivially satisfied by both of its sides being zero, and does not impose any relation between $\Pi_\varepsilon(k)$ and $\Pi_n(k)$. However, integrating it across a shell around k_D , and setting $n(k)$ and all fluxes to zero for $k > k_D$, recovers Eq. 3.

Experimentally, we vary k_D while keeping F_0 fixed and measure $\Pi_n(k_D)$ according to Eq. 2. To mitigate the effects of the long-term few-percent drifts in the initial N , and of the additional atom loss through collisions with the background-gas particles, we perform differential measurements of the cascade-induced atom loss, N_{loss} , with reference measurements taken by setting F_0 to zero in an otherwise identical experimental sequence.

In Fig. 2, we show N_{loss} as a function of the shaking time t_s for various values of U_D . In all cases at short times, we observe no loss (within error). This is consistent with the expectations that no losses occur at $k < k_D \propto \sqrt{U_D}$ and that initially it takes time for the excitations to cascade to k_D , when a steady state can be established (see Fig. 1). For t_s longer than some onset time t_d , the loss rate $\partial N_{\text{loss}}/\partial t$ is essentially constant in time, as long as the total loss is relatively small (<30% of the initial $N \approx 1.2 \times 10^5$). The dashed lines show piecewise linear fits that we use to extract, for each U_D , both t_d and the subsequent initial loss rate, which we identify with the steady-state particle flux

$\Pi_n = \Pi_n(k_D)$. At much longer times, $t_s \gg t_d$, the steady-state assumption can no longer hold, because the losses gradually deplete the low- k source of atoms.

In Fig. 3, we show a log-log plot of Π_n versus U_D (31). We observe power-law behavior with $\Pi_n \propto U_D^{-1.05 \pm 0.08} \propto k_D^{-2.10 \pm 0.16}$. We complement these measurements with numerical simulations based on the Gross-Pitaevskii equation, for the same forcing protocol and without any free parameters [see (27) for details]. The

numerical simulation results are shown by solid circles; a fit to the numerical data gives $\Pi_n \propto U_D^{-1.04 \pm 0.01}$, in good agreement with the experimental data.

The so-called zeroth law of turbulence, first formulated in the context of classical incompressible fluids, stipulates that for fixed forcing, the steady-state rate of energy dissipation tends to a nonzero constant as the viscosity vanishes ($\nu \rightarrow 0$) (5, 33). In our case, this corresponds to keeping F_0 fixed and taking $k_D \rightarrow \infty$ (34). This

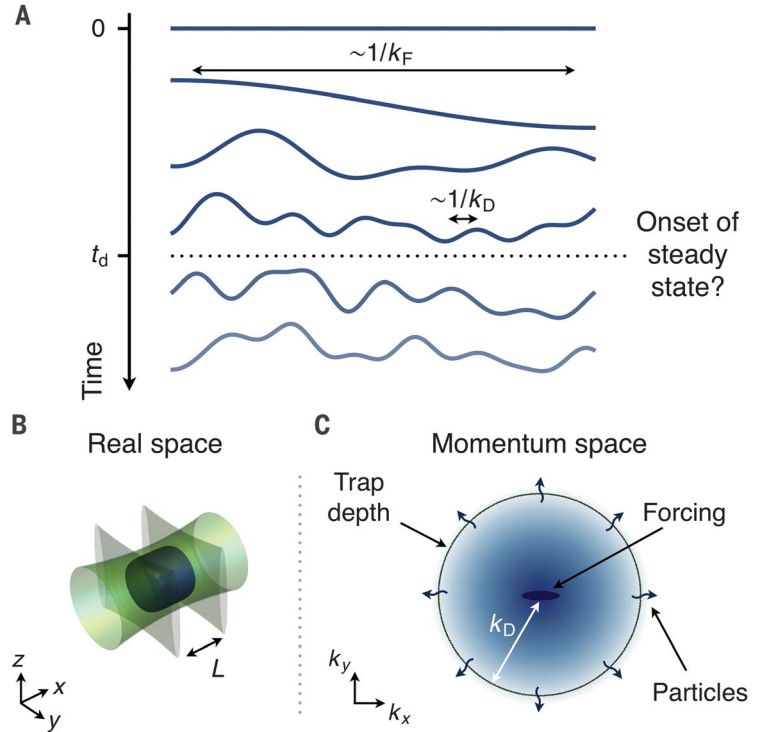
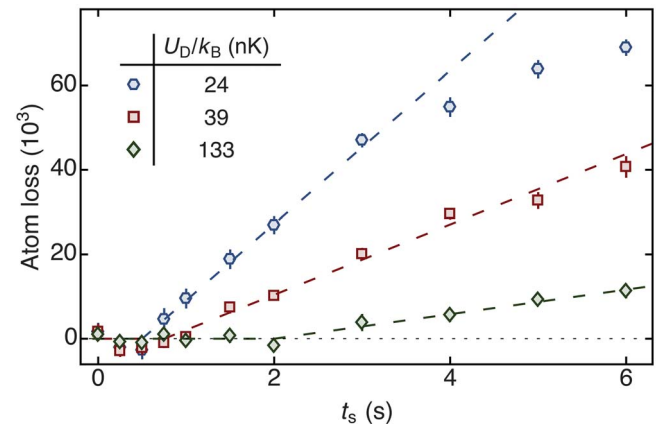


Fig. 1. Turbulent cascade in a box-trapped quantum gas. (A) Cartoon of real-space dynamics of a turbulent wave. Energy is injected by applying a force at a large length scale, $1/k_F$, and propagates to smaller scales because of nonlinear interactions. A steady state can be established when the excitations first reach the small dissipation length scale, $1/k_D$, at time t_d . (B) Sketch of the experimental setting. The atoms are trapped in a finite-depth potential in the shape of a cylindrical box, formed by laser barriers. The shaking force is applied along the x axis. (C) In momentum space, the dissipation scale k_D is set by the trap depth; when excitations propagate to k_D , dissipation occurs in the form of particle loss.

Fig. 2. Atom-loss dynamics associated with the turbulent cascade. The graph shows atoms lost versus shaking time t_s for different trap depths U_D (at $t_s = 0$, the atom number is $N \approx 1.2 \times 10^5$).

Data points show averages of typically 50 measurements. Dashed lines are piecewise linear fits. The systematic uncertainty in U_D values is 20%.



law implies that the particle flux should vanish as $\Pi_n \sim k_D^{-2}$ (see Eq. 3), in excellent agreement with our data. Note that the steady-state energy balance also requires that Π_ε is equal to the rate of energy input into the system, ε . However, energy conservation alone is not sufficient to predict the scaling of Π_n with k_D ,

because it is not a priori obvious that for fixed F_0 the rate at which the system absorbs energy from the drive is not affected by changing k_D (35). Only a posteriori, from Fig. 3 (and the conservation of energy), can we see that in our system the steady-state ε must be independent of the dissipation length scale down to

our lowest k_D . If k_D were changed dynamically, for a system to reach a new steady state the particle flux would have to self-consistently adjust at all $k_F < k < k_D$, because the steady-state Π_n must be both k_D -dependent (to satisfy the zeroth law) and k -independent for a given k_D .

Having established a consistent picture of the steady-state fluxes in our gas, we now turn to the pre-steady-state turbulent dynamics. In Fig. 4A, we depict the early-time dynamics in Fourier space. The forcing, which generates a surplus of atoms at k_F , initiates the cascade at $t_s = 0$. As the cascade front, $k_{cf}(t_s)$, propagates to higher k , the steady-state momentum distribution, $n(k) \propto k^{-\gamma}$, is established in its wake [see also (27)]. The dynamics are dissipationless until k_{cf} reaches k_D (at time t_d); only then is a steady state, with matching ε and $\Pi_\varepsilon(k_D)$, established. Hence, our experimental observations of the initial dissipationless stage of turbulence ($t_s < t_d$), and the dependence of t_d on U_D , give us access to the dynamics of the cascade front in momentum space.

At $t_s < t_d$, the instantaneous particle flux is k -independent for $k < k_{cf}(t_s)$, vanishes for $k > k_{cf}(t_s)$, and must match the rate of the population increase in the inertial range: $n(k_{cf})4\pi k_{cf}^2 dk_{cf} = \Pi_n(k_{cf})dt_s$, so $k_{cf}^{2-\gamma} dk_{cf} \propto \Pi_n(k_{cf})dt_s$. Analogously, for the increase of total energy in the inertial range, $k_{cf}^{4-\gamma} dk_{cf} \propto \Pi_\varepsilon(k_{cf})dt_s$, and $\Pi_\varepsilon(k_{cf})$ is equal to the instantaneous energy injection rate ε .

Assuming that ε , which we found not to depend on k_D in steady state, is also independent of k_{cf} at $t_s < t_d$, then the instantaneous $\Pi_n(t_s)$, at $k < k_{cf}(t_s)$, is $\propto k_{cf}^{-2}$. This gives an elegant unifying picture of the particle fluxes for $t_s < t_d$ and $t_s > t_d$ [Fig. 4A, inset, and (27)]: Π_n is always the same function of the highest k for which the steady-state $n(k)$ has been established (i.e., the lowest k from which there is no backflow), whether that is the instantaneous $k_{cf} < k_D$ (for $t_s < t_d$) or k_D . This self-consistent picture also leads to a quantitative prediction that is verifiable in our experiments: The time independence of ε implies $k_{cf}^{4-\gamma} dk_{cf} \propto dt_s$, which for $\gamma < 5$ and $k_D \gg k_F$ gives a power-law prediction $t_d \propto U_D^\beta$, with $\beta = (5 - \gamma)/2$. Specifically, for our $\gamma = 3.5 \pm 0.1$ (9), we predict $\beta = 0.75 \pm 0.05$.

In Fig. 4B, we show the variation of t_d with U_D . We find that our data are indeed well described by a power law, with $\beta = 0.73 \pm 0.06$, in agreement with our prediction. The results of our numerical simulations (solid circles) show similar behavior with a small systematic offset; a fit to the numerical data gives $\beta = 0.71 \pm 0.01$.

Finally, we note that the criterion for t_d to show scaling behavior, namely $\gamma < 5$ and hence $\beta > 0$, is intimately linked to another important concept in the theory of turbulence. For $\gamma < 5$, the steady-state spectrum has infinite

Fig. 3. Steady-state particle flux. The atom-loss rate Π_n versus the dissipation energy scale U_D (open symbols) is shown on a log-log plot; the three colored points correspond to the data shown in Fig. 2. Solid symbols show the results of numerical simulations (27). The systematic uncertainty in U_D values is 20%. A power-law fit to the experimental data (solid line) gives $\Pi_n \propto U_D^{-1.05 \pm 0.08} \propto k_D^{-2.10 \pm 0.16}$, in agreement with the theoretical prediction.

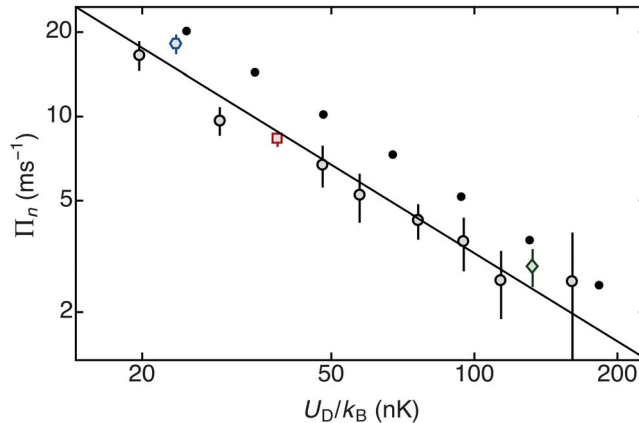
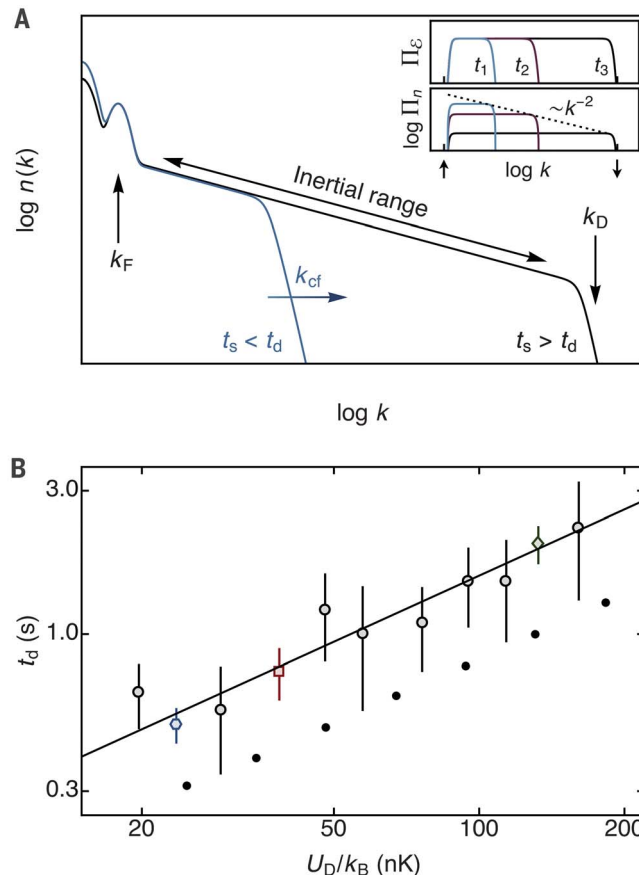


Fig. 4. Establishing the steady state: The cascade-front dynamics.

(A) Momentum-space turbulent dynamics.

Forcing occurs at k_F and the steady-state distribution $n(k)$ is established in the wake of the cascade front $k_{cf}(t_s)$, which propagates outward until it reaches k_D at time t_d . For clarity, here we show an idealized sketch with a very large separation between k_F and k_D ; see (27) for numerical simulations with our experimental parameters. Inset: Consistent picture for the evolution of the energy flux Π_ε and particle flux Π_n for three different times, t_1 (blue) $< t_2$ (purple) $< t_3$ (black), with $t_2 < t_d < t_3$. The forcing and dissipation scales are indicated by the vertical arrows, as in the main panel. (B) Onset time for dissipation. Open symbols show the measured t_d values versus U_D on a log-log plot; the three colored points correspond to the data shown in Fig. 2. Solid symbols show the results of numerical simulations (27). The systematic uncertainty in U_D values is 20%. A power-law fit, $t_d \propto U_D^\beta$, to the experimental data (solid line) gives $\beta = 0.73 \pm 0.06$, in agreement with the prediction $\beta = 0.75 \pm 0.05$.



energy capacity, meaning that it carries infinite energy for $k_D \rightarrow \infty$. It is indeed generally expected for infinite-capacity systems that the cascade front propagates at a finite speed and that the Kolmogorov-Zakharov turbulence spectrum forms behind it (36).

Our work provides a complete, consistent picture of the wave turbulence dynamics at both short (pre-steady-state) and long (steady-state) times. In the broader context of far-from-equilibrium many-body quantum systems, a turbulent quantum gas with a large k_D provides a particularly interesting example of an essentially stationary nonthermal state [see also (37–39)]. The possibility of synthetic dissipation also opens new theoretical perspectives. In the future it would be interesting to engineer arbitrary momentum-cutoff landscapes, which could, for example, allow studies of anisotropic turbulence. By dynamically tuning the dissipation scale or the driving force, it should also be possible to study quenches between different turbulent states.

REFERENCES AND NOTES

1. A. N. Kolmogorov, *Dokl. Akad. Nauk SSSR* **30**, 299–303 (1941).
2. A. Obukhov, *Dokl. Akad. Nauk SSSR* **32**, 22–24 (1941).
3. V. E. Zakharov, V. S. L'vov, G. Falkovich, *Kolmogorov Spectra of Turbulence* (Springer, 1992).
4. S. Dyachenko, A. Newell, A. Pushkarev, V. Zakharov, *Physica D* **57**, 96–160 (1992).
5. U. Frisch, *Turbulence: The Legacy of A. N. Kolmogorov* (Cambridge Univ. Press, 1995).
6. O. Alexandrova *et al.*, *Phys. Rev. Lett.* **103**, 165003 (2009).
7. J. Salort *et al.*, *Phys. Fluids* **22**, 125102 (2010).
8. A. Chepurinov, B. Burkhart, A. Lazarian, S. Stanimirovic, *Astron. J.* **810**, 33 (2015).
9. N. Navon, A. L. Gaunt, R. P. Smith, Z. Hadzibabic, *Nature* **539**, 72–75 (2016).
10. W. D. McComb, *Homogeneous, Isotropic Turbulence: Phenomenology, Renormalization and Statistical Closures* (Oxford Univ. Press, 2014).
11. M. S. Ueberoi, *Phys. Fluids* **6**, 1048 (1963).
12. M. Miyake, M. Donelan, Y. Mitsuta, *J. Geophys. Res.* **75**, 4506–4518 (1970).
13. L. Deike, M. Berhanu, E. Falcon, *Phys. Rev. E* **89**, 023003 (2014).
14. E. A. L. Henn, J. A. Sernan, G. Roati, K. M. F. Magalhães, V. S. Bagnato, *Phys. Rev. Lett.* **103**, 045301 (2009).
15. T. W. Neely *et al.*, *Phys. Rev. Lett.* **111**, 235301 (2013).
16. W. J. Kwon, G. Moon, J. Choi, S. W. Seo, Y. Shin, *Phys. Rev. A* **90**, 063627 (2014).
17. M. C. Tsatsos *et al.*, *Phys. Rep.* **622**, 1–52 (2016).
18. W. J. Kwon, J. H. Kim, S. W. Seo, Y. Shin, *Phys. Rev. Lett.* **117**, 245301 (2016).
19. S. W. Seo, B. Ko, J. H. Kim, Y. Shin, *Sci. Rep.* **7**, 4587 (2017).
20. M. E. Mossman, M. A. Hoefer, K. Julien, P. G. Kevrekidis, P. Engels, *Nat. Commun.* **9**, 4665 (2018).
21. S. P. Johnstone *et al.*, *Science* **364**, 1267–1271 (2019).
22. G. Gauthier *et al.*, *Science* **364**, 1264–1267 (2019).
23. B. V. Svistunov, *J. Mosc. Phys. Soc.* **1**, 373 (1991).
24. D. V. Semikoz, I. I. Tkachev II, *Phys. Rev. Lett.* **74**, 3093–3097 (1995).
25. D. V. Semikoz, I. I. Tkachev, *Phys. Rev. D Part. Fields* **55**, 489–502 (1997).
26. A. L. Gaunt, T. F. Schmidutz, I. Gotlibovych, R. P. Smith, Z. Hadzibabic, *Phys. Rev. Lett.* **110**, 200406 (2013).
27. See supplementary materials.
28. The exponent $\gamma \approx 3.5$ is close to the Kolmogorov-Zakharov prediction for (compressible) weak-wave turbulence in three dimensions, $\gamma = 3$, and in agreement with numerical simulations of the Gross-Pitaevskii equation (9), as well as with a scaling analysis of kinetic equations (29).
29. I. Chantesana, A. P. Orioli, T. Gasenzer, *Phys. Rev. A* **99**, 043620 (2019).
30. O. Cadot, Y. Couder, A. Daerr, S. Douady, A. Tsinober, *Phys. Rev. E* **56**, 427–433 (1997).
31. The lowest U_D we explore corresponds to $U_D/\mu \approx 10$, $U_D/(F_D L) \approx 8$, and $k_D/k_F \approx 23$.
32. Note that it is the total radial flux Π_r , rather than $|\Pi_r|$, that is k -independent in the inertial range.
33. J. C. Vassilicos, *Annu. Rev. Fluid Mech.* **47**, 95–114 (2015).
34. For incompressible flows, the Kolmogorov dissipation length scale, analogous to our $1/k_D$, depends on the viscosity of the fluid (as $\propto \nu^{3/4}$).
35. In most familiar systems, such as the damped harmonic oscillator, the rate at which the system absorbs energy in steady state depends on both the driving and the dissipation.
36. S. Nazarenko, *Wave Turbulence* (Springer, 2011).
37. M. Prüfer *et al.*, *Nature* **563**, 217–220 (2018).
38. C. Eigen *et al.*, *Nature* **563**, 221–224 (2018).
39. S. Erne, R. Bücker, T. Gasenzer, J. Berges, J. Schmiedmayer, *Nature* **563**, 225–229 (2018).
40. N. Navon *et al.*, Research data and code supporting “Synthetic dissipation and cascade fluxes in a turbulent quantum gas”. Apollo (2019).

ACKNOWLEDGMENTS

We thank E. Altman, D. Stamper-Kurn, F. Chevy, and J. Glidden for discussions; T. Hilker for comments on the manuscript; and M. Matsumoto and T. Nishimura for a routine used in the numerical simulations. **Funding:** Supported by EPSRC grants EP/N011759/1 and EP/P009565/1, ERC (QBox), QuantERA (NAQUAS, EPSRC grant EP/R043396/1), AFOSR, and ARO; Trinity College, Cambridge (N.N. and A.L.G.); the David and Lucile Packard Foundation (N.N.); E.U. Marie-Curie program grant MSCA-IF-2015 704832 and Churchill College, Cambridge (R.L.); the Royal Society (R.P.S.); JSPS KAKENHI grant JP16J01683 (K.F.); and JSPS KAKENHI grant 17K05548 and MEXT KAKENHI grant 16H00807 (M.T.). **Author contributions:** N.N. initiated the project. N.N., C.E., and J.Z. collected the data. N.N. and C.E. analyzed the data. K.F. and M.T. performed the numerical simulations. Z.H. supervised the project. All authors contributed extensively to the interpretation of the data and the writing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data that support the findings of this study and the relevant numerical code are available in the Apollo repository (40).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/382/suppl/DC1
Supplementary Text
Figs. S1 to S3
References (41–44)

28 June 2019; accepted 24 September 2019
Published online 3 October 2019
10.1126/science.aau6103



Protein Analysis System

Fluidic Analytics' Fluidity One-W assesses on-target protein interactions in solution, even in crude biological backgrounds such as cell lysates or blood plasma. This capability can produce detailed analyses of proteins in near-native states and in natural environments—facilitating earlier, more accurate analysis of targets for faster decision-making. Fluidic

Analytics has made this possible with its proprietary diffusional sizing technology, which is based on the well-understood relationship between size and diffusion rate, and yields absolute size measurements. The measurements provided by the Fluidity One-W confirm the identity of complexes, allowing scientists to control for off-target binding and false positive measurements, and delivering zero ambiguity for the utmost confidence in results. The instrument's unique abilities and high sensitivity enable researchers to study proteins that are traditionally problematic for other systems, including membrane proteins, multiprotein complexes, and intrinsically disordered proteins.

Fluidic Analytics

For info: +44-(0)-01223-560-432

www.fluidic.com

E. coli Host Cell Protein Kit

Gyros Protein Technologies offers a host cell protein (HCP) kit for automated impurity analysis of biotherapeutics expressed in *Escherichia coli* systems. The kit is optimized for use in Gyrolab systems and has been developed as part of a licensing and supply agreement with Cygnus Technologies (part of Maravai LifeSciences), incorporating their industry standard *E. coli* HCP antibodies and other reagents. The Gyrolab *E. coli* HCP Kit quantifies HCP impurities from *E. coli* expression systems used in the production of biotherapeutics, a critical step in ensuring the efficacy and safety of the drug molecule.

Gyros Protein Technologies

For info: 800-477-6834

www.gyrosproteintechnologies.com

PD-L1 Antibody Panel

In order to give you the best chance of finding the optimal PD-L1 clone for your research assay, you can now get access to four key recombinant monoclonal PD-L1 antibody clones from Abcam (73-10, 28-8, SP142, and CAL10), as well as the anti-mouse recombinant rabbit monoclonal EPR20529 and the mouse monoclonal PDL1/2746. PD-L1 (also known as B7-H1, CD274) is a ligand that binds to PD-1 (programmed cell death-1), expressed on activated T cells, to evade antitumor responses. PD-L1 plays a role in inhibiting T-cell activation and proliferation and has emerged as an important target in cancer treatment. PD-L1 protein detection by immunohistochemistry (IHC) testing is widely used as a predictive biomarker assay for anti-PD-1/PD-L1 therapies. You can

now access all four anti-human recombinant rabbit monoclonal antibody clones in our trial-size panel kit.

Abcam

For info: 888-772-2226

www.abcam.com

Protein Standards

Atlas Antibodies introduces prequantified QPrEST Protein Standards for absolute quantification of proteins in biological samples such as cell lysate and plasma, using LC/MS. QPrESTs consist of 50–150 amino acids identical to a part of the corresponding human protein. They are produced recombinantly in the presence of stable, isotope-labeled (i.e., heavy) amino acids. One advantage of QPrESTs over other available protein standards is that they are added to the sample early in the workflow, prior to proteolytic digestion. The shared amino acid sequence of the QPrEST and the endogenous protein enable the QPrEST to correct for variation introduced during sample digestion, thereby increasing quantification accuracy.

Atlas Antibodies

For info: +46-(0)-8-54-59-58-50

www.atlasantibodies.com

Custom Protein Service

OriGene offers comprehensive, high-quality recombinant custom protein services from different expression systems, in addition to the standard catalog items. Our proprietary protein expression/purification technology maximizes the purity and yield of target proteins in different systems. We offer a complete “start-to-finish” service from the complementary DNA (gene) cloning step to the protein purification step, and downstream modifications such as tag addition/removal, endotoxin testing/removal, conjugations, and any other services you may request. Choose OriGene as your reliable partner for protein-related research, and we can help you accelerate your discovery in a timely, cost-effective, and affordable manner.

OriGene

For info: 888-267-4436

www.origene.com/products/proteins/custom-service

HCP ELISA Detection Kits

AMS Biotechnology has launched a range of ELISA Host Cell Protein (HCP) detection kits that allow you to discover host cell protein impurities in your therapeutic at any phase of product development. The kits are based on a sandwich ELISA test and come in a convenient, 96-well plate format. Specific HCP detection kits are available for popular expression systems, including CHO, *Escherichia coli*, *Pichia*, HEK293, and Protein A. Each kit allows quantification of up to 24 samples. Our HCP detection kits offer 5–10 times more sensitivity, greater antibody coverage, and lower lot-to-lot variations as compared to alternative kits, resulting in a more consistent, reliable antibody supply and a reduction in biomanufacturing costs.

AMS Biotechnology

For info: 800-987-0985

www.amsbio.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.



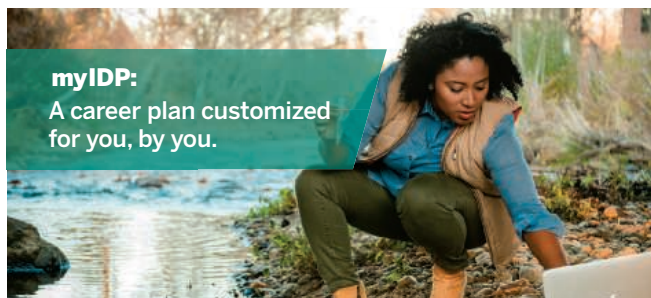
Arts, Sciences
& Education

School of Environment, Arts and Society

Walter and Rosalie Goldberg
Professorship in Tropical
Ecology and Conservation

The Department of Biological Sciences at Florida International University (FIU, <http://biology.fiu.edu>) seeks a top scholar to fill the Walter and Rosalie Professorship in Tropical Ecology and Conservation. We seek qualified individuals whose area of expertise is sustainability, resilience, impact of invasive species, artificial materials, or other aspects of human disruption of terrestrial, aquatic, or marine tropical ecosystems. The ideal candidate will exhibit an interdisciplinary approach to research that will complement existing expertise either within the Department of Biological Sciences or between Biology and allied science departments. The Department of Biological Sciences, (<http://biology.fiu.edu/>) - in the School of the Environment, Arts, and Society (seas.fiu.edu) within the College of Arts, Sciences & Education has ~5,000 majors and 120 graduate students in fields ranging from cell and molecular biology to evolution and ecology. Tropical Research is an area of preeminence at FIU with several institutes and centers relevant to tropical ecology including the International Center for Tropical Botany (icbt.fiu.edu), the Institute of Water and Environment (inwe.fiu.edu), the Tropical Conservation Institute (tci.fiu.edu), the Southeast Environmental Research Center (serc.fiu.edu), The Sea Level Solutions Center (slsc.fiu.edu), The Center for Coastal Oceans Research (ccor.fiu.edu), and the International Hurricane Research Center (www.ihr.fiu.edu). The successful candidate will be expected to have an exceptional publication record, bring and maintain an active externally funded research program, participate in the Biology graduate program, as well as teach a course in their area of expertise. The minimum requirements are a doctoral degree from an accredited institution, associate or full professor or equivalent status, and demonstrated record of achievement in academic research, teaching and service.

Qualified candidates are encouraged to apply to **Job Opening ID 519325** at facultycareers.fiu.edu and attach a cover letter, curriculum vitae, and statements of research, teaching, and service as a single pdf file. Please also provide names and contact information for at least three references who will be contacted upon submission of application. To receive full consideration, applications and required materials should be received by **November 18, 2019**. Review will continue until position is filled. Please direct inquiries about this search to the chairman of the search committee, Dr. Joel Trexler trexlerj@fiu.edu.



myIDP:

A career plan customized
for you, by you.

For your career in science, there's only one **Science**

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.



Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers
N/AAS

In partnership with:



Marine Biological Laboratory



Marine Biological Laboratory 2020 Whitman Center Fellowships



THE MARINE BIOLOGICAL LABORATORY, a hub for research and education and an affiliate of the University of Chicago, convenes biologists from around the world each year to advance the mission of biological discovery. We are now accepting applications for **Whitman Center Fellowships** for 2020. Support is available for scientists to come to the Marine Biological Laboratory for 4 to 10 weeks to conduct research, year-round.

We particularly encourage applications from individuals or collaborative groups focused on the following:

- Evolutionary, genetic, and genomic approaches in regenerative and developmental biology, microbiomes, and neuroscience with an emphasis on marine organisms
- Integrated imaging and computational approaches to illuminate cellular function and biology emerging from the study of marine and other organisms
- Integrated approaches to the study of microbial communities and marine organisms in coastal communities.

The 2020 Whitman Center Fellowships include **Whitman Center Early Career Fellowships**, specifically designated for individuals less than 10 years from their doctoral degree who wish to focus on these areas of biological discovery, and the **Whitman Center E.E. Just Fellowships** for underrepresented minorities in science or individuals from HBCUs or MSIs.

Whitman Center Fellowships cover laboratory rental and housing costs. The MBL offers access to state-of-the-art instrumentation, innovative imaging technology, DNA sequencing and informatics, year-round availability of model freshwater and marine organisms, and modern laboratory facilities.

The Marine Biological Laboratory hosts more than 1,000 researchers, postdocs, and graduate students from around the world to participate in scientific discovery courses, research, lectures, and field studies. As a convener of biology, the Marine Biological Laboratory is well known for fostering a highly collaborative environment, with scientists and students engaged in intensive research in a collegial and informal atmosphere.

Eligible applicants must hold appointments at accredited universities, colleges or research institutions anywhere in the world. While applications will be evaluated on the basis of scientific merit, we are especially interested in individuals from diverse backgrounds, experiences and perspectives, especially those under represented in science.

mbl.edu/research/whitman-fellowships

research@mbl.edu

Application Deadline: **December 15**



Tenure-Track Assistant Professor Faculty Position in Cellular or Molecular Immunology

The Department of Biological Sciences in the College of Natural Sciences and Mathematics at the University of Toledo seeks to fill a **tenure-track assistant professor** faculty position. Appointment at a higher rank will also be considered. Applicants are required to hold a Ph.D. with post-doctoral experience in Biological Sciences or a related field. Candidates in cellular or molecular aspects of **immunology** will be given priority. However, outstanding candidates in other areas of cellular or molecular biology will also be considered. It is important that candidates are able to collaborate with other faculty within the Department. Innate and other host disease defense mechanisms are part of our Departmental core research strengths, in addition to Cell Architecture and Dynamics, cell division/cancer biology, neuroscience, and regulation of gene expression, unified by a common interest in cell signaling. Other Departments in the College of Medicine on the near-by Health Science Campus of the University complement these areas. Facilities include a modern research complex with sophisticated laboratories and access to outstanding research instrumentation including, flow cytometers/cell sorters, confocal and super-resolution microscopes, and a new transmission electron microscope. The successful candidate is expected to advance excellence in the Department of Biological Sciences at the University of Toledo by establishing an externally-funded research program and contributing to undergraduate and graduate education. The Department offers B.A., B.S., M.S., and Ph.D. degrees.

Additional information is available on the Departmental website: www.utoledo.edu/nsm/bio/index.html. Salary and start-up funds are competitive. Review of applications will begin immediately and continue until the position is filled. The starting date for this position will be August 2020. Interested candidates must apply at <https://jobs.utoledo.edu> with a letter of application, curriculum vitae, statements of research and teaching interests, and the names and contact information for three referees. Email inquiries may be directed to Lirim Shemshedini at LIRIM.SHEMSHEDINI@utoledo.edu or Scott.Leisner@utoledo.edu.

The University of Toledo is an Affirmative Action Equal Opportunity Employer. All qualified applicants will receive consideration of employment without regard to race, color, religion, sex, national origin, disability, or protected Vets status. Qualified women and minorities are especially encouraged to apply. In addition, we are particularly interested in receiving applications whose research focuses on underrepresented populations and/or who have experience mentoring underrepresented minority students. The University of Toledo is committed to achieving inclusive excellence through a diverse workforce.



Rutgers-The State University of New Jersey seeks applicants for appointment at the rank of Assistant Professor/Associate Professor/Professor, Basic Science in the Department of Pathology, Immunology and Laboratory Medicine at the Rutgers-New Jersey Medical School located in Newark, New Jersey. This is a full-time Assistant /Associate or full Professor position. This position will support the research mission within the department, as well as the School. The research areas include Infectious Diseases, Cancer and Immunology. The successful candidate is expected to develop a strong independent research program within the department. In addition, he/she will participate in education and activities supporting research and scholarship. The successful candidate will bring in extramural funding with research interests and strength in Infectious Diseases, Immunology, and Cancer to complement and further the Department's research mission. The successful candidate will initiate translational research in collaboration with the clinical faculty and participate in teaching of the RBHS-Graduate School of Biomedical Sciences (GSBS). Minimum education and experience: MD, PhD, or MD/PhD. Extramural funding is expected in the research area(s) of infectious Diseases, Immunology and Cancer. Teaching experience. Applications or inquiry should be submitted to the following email address: chen.liu@rutgers.edu.

Rutgers University is an AA/EEO Employer.

SOMETIMES THE GRASS REALLY IS GREENER SOMEPLACE NEW.



Find your next job at ScienceCareers.org

There's scientific proof that when you're happy with what you do, you're better at what you do. Access career opportunities, see who's hiring and take advantage of our proprietary career-search tools. Get tailored job alerts, post your resume and manage your applications all in one place: sciencecareers.org

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS



Statement required by the Act of 12 August 1970, Section 3685, Title 39, United States Code, showing the ownership, management, and circulation of:

1–9. *Science*, Publication No. 0036-8075, is published weekly on Friday, except the last week in December, at 1200 New York Avenue, N.W., Washington, DC 20005. Date of filing: 02 October 2019. This is also the address of the publisher, the editor, and the managing editor, who are, respectively, Bill Moran, Holden Thorp, and Monica M. Bradford.

10. The owner is the American Association for the Advancement of Science, 1200 New York Avenue, N.W., Washington, DC 20005. Stockholders: None.

11. Known bondholders, mortgages, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities: None.

12. The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding 12 months.

13–15. The average number of copies of each issue during the preceding 12 months is (A) Total number of copies printed: 62,943; (B) Paid circulation: 58,402; (1) Paid/Requested outside-county mail subscriptions stated on form 3541: 51,244; (2) Paid/Requested in-county subscriptions stated on form 3541: 0; (3) Sales through dealers and carriers, street vendors, counter sales: 7,151. (4) Other classes mailed through USPS: 7; (C) Total paid circulation: 58,402; (D) Free distribution: samples, complimentary, and other free copies: 3,352; (1) Outside-county as stated on form 3541: 2713; (2) In-county as stated on form 3541: 0; (3) Other classes mailed through the USPS: 2; (4) Free distribution outside of mail carrier or other means: 637; (E) Total free distribution: 3,352; (F) Total distribution: 61,754; (G) Copies not distributed: 1,190; (H) Total: 62,944; (I) Percent paid and/or Requested Circulation: 94.6%.

Actual number of copies of single issue (9/20/2019) published nearest to filing date are (A) Total number of copies printed: 60,867; (B) Paid circulation: 56,686; (1) Paid/Requested outside-county mail subscriptions stated on form 3541: 49,906; (2) Paid/Requested in-county subscriptions stated on form 3541: 0; (3) Sales through dealers and carriers, street vendors, counter sales: 6,774; (4) Other classes mailed through USPS: 6; (C) Total paid circulation: 56,686; (D) Free distribution: Samples, complimentary, and other free copies: 3,181; (1) Outside-county as stated on form 3541: 2,749; (2) In-county as stated on form 3541: 0; (3) Other classes mailed through the USPS: 2; (4) Free distribution outside of mail: Carrier or other means: 430; (E) Total free distribution: 3,181; (F) Total distribution: 59,867; (G) Copies not distributed: 1,000; (H) Total: 60,867; (I) Percent paid and/or Requested Circulation: 94.7%.

I certify that the statements made above are correct and complete. (signed) Bill Moran, Publisher.

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

Follow us for jobs,
career advice & more!



@ScienceCareers



/ScienceCareers



Science Careers

ScienceCareers.org

RWTH AACHEN
UNIVERSITY

JARA | Jülich Aachen
Research
Alliance

JÜLICH
Forschungszentrum

Focusing Expertise – Shaping the Future: The Jülich Aachen Research Alliance (JARA) is an innovative cooperation model between RWTH Aachen University and Forschungszentrum Jülich. This Alliance brings together an internationally respected university of science and technology and one of the leading national research centres in Europe.

As a member of the Helmholtz Association, Forschungszentrum Jülich makes an effective contribution to solve major challenges facing society in the fields of information, energy, and bioeconomy. They focus on varied tasks in the area of research management and utilize large, often unique, scientific infrastructures. Come and work with around 6,100 colleagues across a range of topics and disciplines at one of Europe's largest research centres.

As one of Germany's Universities of Excellence, RWTH enjoys a global reputation for high-quality teaching and research. The university seeks to drive the convergence of knowledge, methods and findings from its different research fields and to integrate this in-depth knowledge into its interdisciplinary profile areas. RWTH's dynamic and creative environment is characterized by efficient networks, institutionalized collaboration and the strong innovative capacity of RWTH Aachen Campus, one of the largest technology-oriented research landscapes in Europe.

The research performed at the Peter Grünberg Institute (PGI) ranges from physical concepts and emerging materials to novel nanoelectronic devices. It is an active partner in the Cluster of Excellence "Matter and Light for Quantum Computing" (ML4Q; cooperation by the universities of Aachen, Cologne, and Bonn, as well as the Research Center Jülich) and the European "Quantum Flagship". Supported by the Jülich Supercomputing Centre (JSC) and the Helmholtz Nanoelectronic Facility (HNF), a 1000 m² clean room, PGI will further strengthen its activities on quantum information. It anticipates the next level by harbouring the Helmholtz Quantum Center (HQC), home of the newly founded institute on quantum computing. In this context RWTH Aachen University and Forschungszentrum Jülich are jointly seeking a

Director for the PGI – Institute for Quantum Computing (PGI-13) appointed as Professor for Quantum Computing (W3, Jülicher Modell) at RWTH Aachen University, Department of Physics

In close cooperation with other national and international groups we are initiating a center of expertise on fundamentals of quantum computing. Within this effort the goal of the new institute is the development of experimental quantum processors. The focus should be on multi-qubit circuits using scalable solid state qubits. Areas of interest include the coupling of qubits, high fidelity and scalable qubit control, as well as system integration. The institute should seek strong synergies with existing initiatives within JARA, the planned Helmholtz Quantum Center and the Cluster of Excellence ML4Q.

Teaching requirements (2 hours/week) could, for example, be fulfilled with courses in quantum science and technology within the physics programs at RWTH Aachen. A doctoral degree and an exemplary record of research achievements at the group leader level are required.

RWTH Aachen University and Forschungszentrum Jülich are equal opportunities employers and pursue a policy of excluding all types of discrimination. Applications from women scientists are particularly welcome. Applications from women will be given preference in the case of equal suitability, qualifications and experience, unless special reasons concerning the person of a male candidate outweigh these considerations. Both institutions also are family- and disability-friendly and offer support for dual career couples.

Applications should be in English and should be sent with the usual documentation (CV, copies of certificates, list of publications, teaching experience, brief summary of previous research activities including a list of third-party funding, research concept for the advertised position), by the 15-12-2019 to (e-mail preferred):

**Board of Directors of
Forschungszentrum Jülich GmbH**
D-52425 Jülich
berufungen@fz-juelich.de

and **Dean of the Faculty of Mathematics,
Computer Science and Natural Sciences at
RWTH Aachen University, Prof. U. Simon**
D-52056 Aachen
application@fb1.rwth-aachen.de

Further information can be found at: www.jara.org/jara-fit/PGI-13



By Nathan Pickle

Escaping ‘The Waiting Place’

In Dr. Seuss’s classic book *Oh, the Places You’ll Go!* he warns us that sometimes in life we may “grind on for miles across weirdish wild space, headed, I fear, toward a most useless place. The Waiting Place.” Little did I know it, but after I emerged from the “weirdish wild space” of graduate school, The Waiting Place was exactly where I was headed. I had accepted a postdoc position to develop new skills while embarking on what I anticipated would be a quick, productive search for an industry job. Instead that search dragged on, filled with false hopes and disappointments. I often felt I was going nowhere. But now that I’ve passed the 1-year mark in my first position outside of academia, I can look back and say that I learned a valuable skill: the ability to wait actively. Here are three strategies I found most helpful.

SEEK GUIDANCE. As a freshly minted Ph.D., I had plenty of material for my resume, but I wasn’t sure how to pitch my skills to potential employers. To address this, I sought out a career coach who specialized in helping people with Ph.D.s in science and engineering. Hiring a professional consultant wasn’t cheap, but I viewed it as an investment in another valuable skill: marketing myself. His guidance not only improved my resumes, but also gave me a better general understanding of how to present myself in meetings with recruiters and in interviews. I also took advantage of many free opportunities for career advice, including university-sponsored career forums, skills assessments with an on-campus career counselor, and numerous online resources for job seekers.

MEET NEW PEOPLE. At first, I felt that I didn’t know “the right people” to help me get a job, especially because I was looking for an industry position and my contacts were mostly in the academic world. However, once I started to look, I was surprised to realize how many second-degree connections—friends of friends—I actually had in industry. I reached out to many of them and asked to talk about their work. I also asked friends and colleagues for introductions to people in their networks. And I found ways to meet people outside my niche research area, for example by attending presentations by startup companies, meetings of technology interest groups, and on-campus socials.

It sometimes felt awkward to put myself out there, but networking helped me learn about many companies and



**“As a freshly minted Ph.D. ...
I wasn’t sure how to pitch my skills
to potential employers.”**

industries. It also led to job interviews, including with a Silicon Valley tech giant, a premier children’s hospital, and a highly ranked liberal arts college. Organizations like to hire people they can trust, so having someone on the inside vouch for you can help you rise to the top of the applicant stack and get an interview.

PLAY THE LONG GAME. In the end, I got a job I didn’t even apply for. During my first year as a postdoc, I applied for an opportunity I found on an online job board. This led to a promising email exchange, and I had a very positive interaction with one of the company’s researchers at a conference. Unfortunately, the opportunity fell through. Another long year passed. I followed up with the company once or twice, but mostly I pursued other leads. Then, out of the blue, the researcher I had met contacted me about a new opening. It turned out I was a perfect fit for the company’s current projects, and after a couple short phone calls I had an offer. It took almost 2 years to come to fruition, but the effort I had invested in building and maintaining a relationship with the company ultimately paid off.

The Waiting Place was frustrating, but I learned it doesn’t have to be useless. Perseverance pays off. As Dr. Seuss reminds us: “And will you succeed? Yes! You will, indeed! (98 and 3/4 percent guaranteed.)” ■

Nathan Pickle is a research scientist at CFD Research Corporation in Huntsville, Alabama. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER